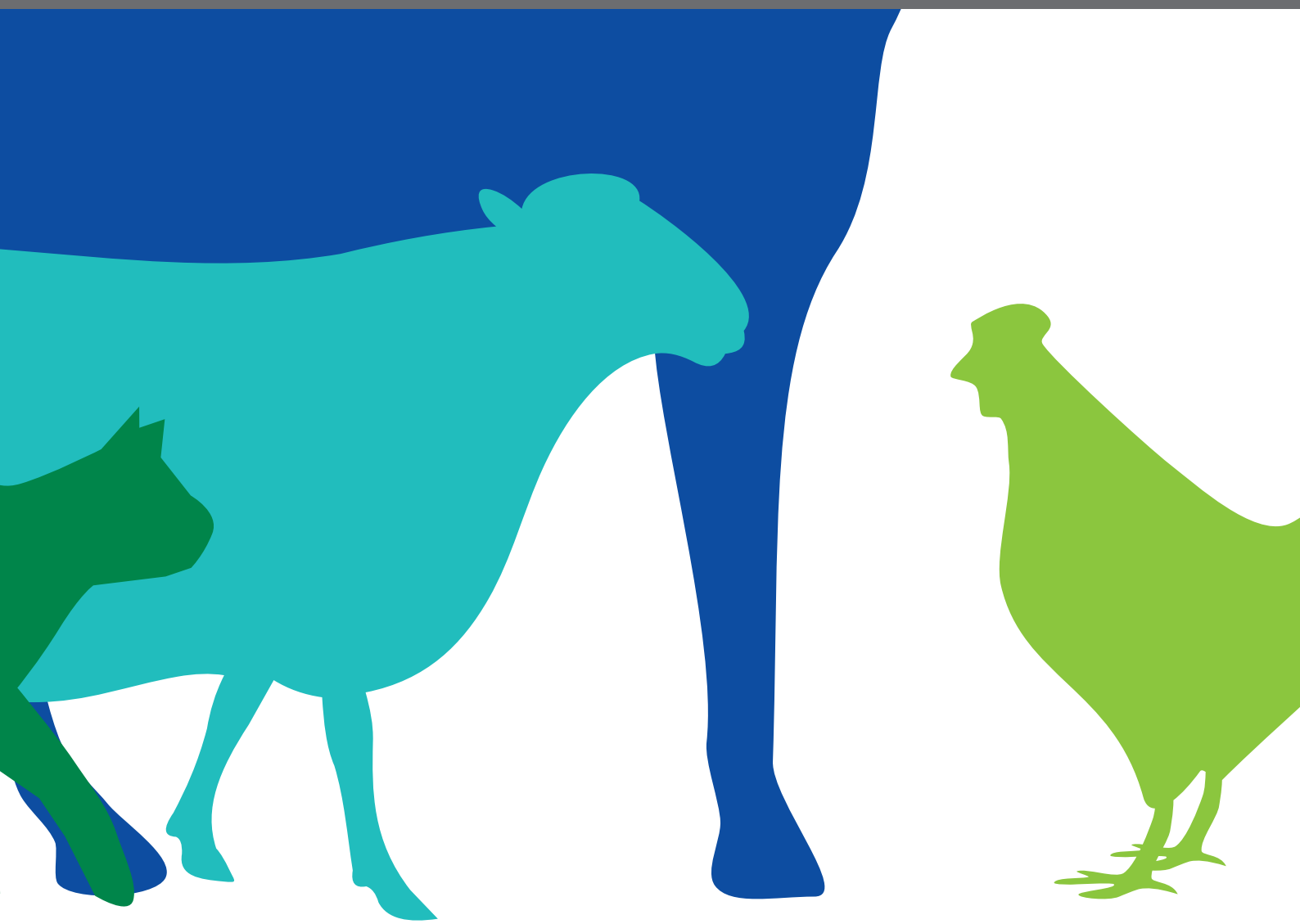
PRINCIPLES AND CHALLENGES OF FUNDAMENTAL METHODS IN VETERINARY EPIDEMIOLOGY AND ECONOMICS

EDITED BY: Salome Dürr, Victoria J. Brookes and Andres M. Perez
PUBLISHED IN: Frontiers in Veterinary Science



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-152-9
DOI 10.3389/978-2-88971-152-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

PRINCIPLES AND CHALLENGES OF FUNDAMENTAL METHODS IN VETERINARY EPIDEMIOLOGY AND ECONOMICS

Topic Editors:
Salome Dürr, University of Bern, Switzerland
Victoria J. Brookes, Charles Sturt University, Australia
Andres M. Perez, University of Minnesota Twin Cities, United States

Citation: Dürr, S., Brookes, V. J., Perez, A. M., eds. (2021). Principles and Challenges of Fundamental Methods in Veterinary Epidemiology and Economics. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-152-9

Table of Contents

frontiers in Veterinary Science

EDITORIAL
published: 14 June 2021
doi: 10.3389/fvets.2021.705980

Editorial: Principles and Challenges of Fundamental Methods in Veterinary Epidemiology and Economics

Salome Dürr[1]*, *Victoria J. Brookes*[2,3] *and Andres M. Perez*[4]

[1] Veterinary Public Health Institute, Vetsuisse Faculty, University of Bern, Bern, Switzerland, [2] Faculty of Science, School of Animal and Veterinary Sciences, Charles Sturt University, Wagga Wagga, NSW, Australia, [3] Graham Centre for Agricultural Innovation (NSW Department of Primary Industries and Charles Sturt University), Wagga Wagga, NSW, Australia, [4] College of Veterinary Medicine, University of Minnesota, St Paul, MN, United States

Keywords: method, guideline, introduction, discipline, epidemiology, economics

Editorial on the Research Topic

Principles and Challenges of Fundamental Methods in Veterinary Epidemiology and Economics

The discipline of veterinary epidemiology focuses on the investigation of the dynamics, frequency, and determinants of diseases in populations of veterinary interest. Epidemiological methods are continuously changing, as new tools and techniques become available, often borrowed from other disciplines and adapted to veterinary science objectives. Therefore, there is a need for epidemiologists to become acquainted with both existing and emerging methods to take advantage of all available approaches to support the prevention and control of disease in animal populations.

Internationally recognised epidemiologists have been invited to contribute to this Research Topic with articles that cover existing and emerging areas of epidemiological research with a focus on their use in veterinary research. Whilst some methods presented here have been used for decades, others have evolved relatively recently. To promote their use by the broad community of veterinary epidemiologists worldwide, the articles have been written to introduce methodologies to researchers who are relatively new to the respective topic. The principles, advantages, challenges and limitations, as well as perspectives on how these methods will evolve given their use in a veterinary epidemiology context, are discussed in each article, so that they can be used as guidelines for application. To support the practical use of the methods presented in the articles, code and example data have been provided where possible.

Articles have been grouped into three sections: (1) assessing the literature, collecting data, and measuring disease, (2) identifying epidemiological associations and exploring disease patterns, and (3) modelling disease and estimating its economic impact.

Methods to assess the literature, collect data, and measure the impact of disease are fundamental aspects of veterinary epidemiology and economics. Six articles are allocated to this area in this Research Topic. Sargeant and O'Connor provided an overview on scoping review, systematic reviews, and meta-analysis, and their application in veterinary science. They highlight that these methods are becoming increasingly popular for both researchers and practitioners; understanding the distinction between review types is important to be fit for purpose. For example, whilst scoping reviews might map the broad knowledge around an area of veterinary science, a systematic review might be more useful to identify literature relevant to a more specific area. Finally, meta-analyses quantitatively combine the results from multiple studies that have been

OPEN ACCESS

Edited and reviewed by:
Flavie Vial,
Animal and Plant Health Agency, United Kingdom

***Correspondence:**
Salome Dürr
salome.duerr@vetsuisse.unibe.ch

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *06 May 2021*
Accepted: *21 May 2021*
Published: *14 June 2021*

Citation:
Dürr S, Brookes VJ and Perez AM (2021) Editorial: Principles and Challenges of Fundamental Methods in Veterinary Epidemiology and Economics. Front. Vet. Sci. 8:705980. doi: 10.3389/fvets.2021.705980

identified by systematic reviews. Hu et al. describe a type of meta-analysis, the Bayesian network meta-analysis, in more detail and illustrate the procedures and stepwise workflow, including a description of how to informatively present the results in ranking plots and treatment risk posterior distribution plots. Brennan et al. present another approach for systematic evidence synthesis: Critically Appraised Topics (CATs). This method uses the same principles as a systematic review, but is aimed at addressing a clinically-based question from a veterinary professional to support evidence-based clinical practice. The authors illustrate five steps of CATs based on an example, and emphasise the clinical relevance and practicalities. Stevenson provides a guide to a fundamental concept in epidemiology: sample size estimation. Justification of the number of subjects enrolled into a study and how this has been calculated are a core requirement of any epidemiological study, to demonstrate sufficient power whilst balancing resources such as time and sampling cost. Animals are typically aggregated into groups leading to a lack of independence of observations, and approaches are discussed to overcome this issue. The article by Degeling and Rock presents principles of qualitative research for One Health projects. They highlight the potential of collaborative projects between qualitative researchers and veterinary epidemiologists by emphasizing how qualitative research can contribute to better interpretation of findings in different cultural, economic, historical, and social contexts. As such, qualitative methods support epidemiological researchers to develop policy so that it can be implemented with a more politically and socio-culturally meaningful approach. Several methods, such as interviews, participant observations and working with groups are presented, as well as useful ways to analyze such data. Alders et al. contribute an article on participatory epidemiology (PE) in veterinary science, a method that has evolved to embrace knowledge, experience, and motivations of relevant stakeholders, such as animal caretakers and owners, for identification and assessment of animal disease problems. The review article describes the evolution of PE, its philosophy and principles for effective application, and the importance of data triangulation and gender- and minority-sensitive approaches.

Identifying epidemiological associations and exploring spatial and temporal disease patterns are key domains for veterinary epidemiology and economics. Five articles were collated in this section. Kratzer et al. introduce the readers to Bayesian network (BN) modelling, described as a flexible analytical framework for complex epidemiological datasets. In veterinary science, we are often confronted with datasets containing interdependent variables, which challenge classical uni- and multi-variable regression models used for risk factor analyses. BN modelling is an approach which aims to overcome these issues, and untangle direct and indirect relationships between variables. BN modelling is described and applied using a stepwise approach to a veterinary dataset, and results are compared to a classical regression approach. Kanankege et al. illustrate an overview of spatiotemporal visualisation and analytical tools (SATs) in population-level eco-epidemiological research, and present them in a framework for choosing the appropriate method for a specific research question and dataset. Following a stepwise process based on six research questions, researchers are directed to select a suitable SAT, belonging to one of four categories: (1) visualisation and descriptive analysis, (2) spatial or spatiotemporal dependence and pattern recognition, and (3) spatial smoothing and interpolation and (4) geographic correlation studies for testing inferences in spatial dependent datasets. Escobar presents an article on ecological niche modelling. He highlights the increasing importance of distributional ecology in the field of epidemiology, since the biotic interactions between pathogens and hosts are crucial for infectious disease diversity, distribution and maintenance. Development of interdisciplinary research methods that bring together ecology and epidemiology, such as ecological niche modelling, will improve predictions for infectious disease abundance. Ward et al. present and discuss methods used to analyse time-series data in veterinary science, focusing on ARIMA (Autoregressive Integrated Moving Average) models. Time-series datasets are relatively common in animal disease monitoring and surveillance systems, but can also originate from animal production or welfare datasets that are increasingly available in modern animal production systems. Although many datasets in veterinary science could be analyzed using this method, it was found to be relatively rarely used in this field. The stepwise instructions and code provide veterinary epidemiologists with foundation skills in this method. Alkhamis et al. introduced the use of phylodynamic methods as a recently evolved method for emerging and endemic animal viral disease surveillance. Phylodynamic methods offer the possibility to integrate spatio-temporal epidemiology and evolutionary dynamics into one analysis framework by using single Bayesian statistical techniques on the phylogeny of viruses in populations. The practical steps required to perform phylodynamic analyses (sequence preparation, preliminary phylogenetic analysis, selecting and running phylodynamic models, and visualisation of the model outcomes) are presented, and the robustness of different models is tested. Challenges for integrating phylodynamic methods in routine animal disease surveillance activities are also discussed.

The use of infectious disease modelling has been increasing over previous decades and nowadays comprises diverse methodologies. Three articles describe different methods, and a fourth focuses on the estimation of the economic impact of infectious diseases. Brzoska et al. present a relatively new method called stochastic block modelling (SBM) to unravel complex networks such as those existing in animal trade. SBM splits such networks into smaller units of nodes (for example, farms) that show similar network properties. The method was shown to perform better for informing trade restrictions to control diseases, compared to the more established community detection method or trade restrictions based on geographical boundaries. Kinsley et al. also illustrate novel network methodologies, namely multilayer and multiplex networks. Multilayer networks account for different modes of spread of pathogens between hosts, and therefore, consider multiple layers of contacts resulting in multiplex networks. Also, different types of hosts can harbor the same pathogens within one ecosystem, which can be illustrated by interconnected multilayer networks. In

this article, these techniques are reviewed, and applied to an example to demonstrate how these models can capture disease dynamics in complex host-pathogen systems. Kirkeby et al. present an introduction to mechanistic modelling of disease transmission, focusing on individual-based models that allow inclusion of heterogeneity between individual epidemiological units. The article illustrates and describes important steps before, during and after model programming. Model verification, validation, convergence analysis, and sensitivity analysis are described and discussed, and examples provided for each of these steps. Gethmann et al. present gross margin analysis (GMA) to estimate the economic impact of an infectious disease outbreak. They implement GMA within the Excel add-in @Risk (Palisade), an easy-to-use, commercial tool for stochastic simulations. Within their @Risk model, direct costs (for example, production losses, animal deaths, veterinary treatment) and indirect costs (for example, surveillance, measures for animal export, disease control, vector monitoring, and administration) can be individually entered. Such tools can therefore be suitable for economic impact estimation, thus using GMA for broad applications in veterinary economic.

In summary, this Research Topic provides an overview of existing, emergent and novel methods with applications in veterinary epidemiology. We expect that this contribution will provide the veterinary epidemiology community with resources to improve their knowledge on existing and novel methods, thus supporting the prevention, surveillance and control of diseases that impact both human and animal health.

AUTHOR CONTRIBUTIONS

All the authors were involved in writing the article and agreed to the final version of this editorial.

ACKNOWLEDGMENTS

We thank all authors, additional editors and reviewers for the contribution to the Research Topic.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

METHODS
published: 28 January 2020
doi: 10.3389/fvets.2020.00011

Scoping Reviews, Systematic Reviews, and Meta-Analysis: Applications in Veterinary Medicine

Jan M. Sargeant[1]* *and Annette M. O'Connor*[2]

[1] Department of Population Medicine, University of Guelph, Guelph, ON, Canada, [2] Department of Veterinary Diagnostic and Production Animal Medicine, Iowa State University College of Veterinary Medicine, Ames, IA, United States

Evidence-based decision making is a hallmark of effective veterinary clinical practice. Scoping reviews, systematic reviews, and meta-analyses all are methods intended to provide transparent and replicable ways of summarizing a body of research to address an important clinical or public health issue. As these methods increasingly are being used by researchers and read by practitioners, it is important to understand the distinction between these techniques and to understand what research questions they can, and cannot, address. This review provides an overview of scoping reviews, systematic reviews, and meta-analysis, including a discussion of the method and uses. A sample dataset and coding to conduct a simple meta-analysis in the statistical program R also are provided. Scoping reviews are a descriptive approach, designed to chart the literature around a particular topic. The approach involves an extensive literature search, following by a structured mapping, or charting, of the literature. The results of scoping reviews can help to inform future research by identifying gaps in the existing literature and also can be used to identify areas where there may be a sufficient depth of literature to warrant a systematic review. Systematic reviews are intended to address a specific question by identifying and summarizing all of the available research that has addressed the review question. Questions types that can be addressed by a systematic review include prevalence/incidence questions, and questions related to etiology, intervention efficacy, and diagnostic test accuracy. The systematic review process follows structured steps with multiple reviewers working in parallel to reduce the potential for bias. An extensive literature search is undertaken and, for each relevant study identified by the search, a formal extraction of data, including the effect size, and assessment of the risk of bias is performed. The results from multiple studies can be combined using meta-analysis. Meta-analysis provides a summary effect size, and allows heterogeneity of effect among studies to be quantified and explored. These evidence synthesis approaches can provide scientific input to evidence-based clinical decision-making for veterinarians and regulatory bodies, and also can be useful for identifying gaps in the literature to enhance the efficiency of future research in a topic area.

Keywords: evidence synthesis, scoping reviews, systematic reviews, meta-analysis, veterinary

OPEN ACCESS

Edited by:
Andres M. Perez, University of Minnesota Twin Cities, United States

Reviewed by:
Katharina D. C. Stärk, Federal Food Safety and Veterinary Office, Switzerland
Javier Sanchez, University of Prince Edward Island, Canada

***Correspondence:**
Jan M. Sargeant
sargeanj@uoguelph.ca

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *14 October 2019*
Accepted: *08 January 2020*
Published: *28 January 2020*

Citation:
Sargeant JM and O'Connor AM (2020) Scoping Reviews, Systematic Reviews, and Meta-Analysis: Applications in Veterinary Medicine. Front. Vet. Sci. 7:11. doi: 10.3389/fvets.2020.00011

BACKGROUND

Evidence-based decision-making is a hallmark of veterinary clinical practice and veterinary public health. Evidence-based veterinary medicine has evolved from principles of evidence-based medicine developed in the human healthcare literature. The evidence-based medicine approach integrates patient values, clinical expertise, and scientific evidence to make decisions about the clinical care of patients (1, 2). Within this approach, scientific evidence is derived from the results of research studies. However, clinical trials may differ in their inclusion criteria and recruitment, and trials are conducted on a sample of the target population; therefore, the results of a single study represent a random result from a distribution of possible trial results (3, 4). Additionally, there is empirical evidence that the first study on a given topic will have the largest effect size, with diminishing or contradictory effect sizes reported in subsequent studies (3, 5). As a consequence of these concepts, decision-makers should use the body of evidence rather than a single study result, as the unit of concern for making evidence-based decisions. However, it is time consuming for veterinarians and others involved in veterinary decision-making to identify, acquire, appraise, and apply the available literature on a given topic. For instance, a simple search in PubMed using the search string cattle AND (BRD or "bovine respiratory disease") AND (vaccine or vaccination) resulted in the identification of 286 potentially relevant articles (search conducted Jan 10th 2020). Thus, it is essential both to replicate research and to have a means of combining (synthesizing) the results of multiple studies addressing the same research question.

Evidence synthesis refers to the combination of results from multiple sources. There is a plethora of methodologies for undertaking evidence synthesis for various types of information or types of synthesis questions (6). This paper focuses on two common evidence synthesis tools used in veterinary medicine: scoping reviews and systematic reviews. Meta-analysis, the statistical summarization of results from multiple studies, is the analytical component of a systematic review which can be undertaken when there is a sufficient body of literature identified in the review. Both scoping reviews and systematic reviews are methods to synthesize existing literature by following a series of structured and documented steps, and using methods intended to reduce the risk of bias. However, the two types of reviews answer different research questions. Scoping reviews are a descriptive study design, intended to chart or map the available literature on a given topic. By contrast, systematic reviews answer a specific question, often related to clinical decision-making, with the ideal end product being a summarized effect or effect size across multiple studies or an exploration of sources of heterogeneity (differences among studies in the effect or effect size).

METHODS

Scoping Reviews

Scoping reviews are used to describe the available literature on a topic (often referred to as charting or mapping). The specific objectives of a scoping review might be to describe the volume and nature of the existing literature in a topic area, to determine the feasibility of conducting a systematic review for a specific review question within a topic area, or to identify gaps in the body of literature on a topic (7, 8). The approach was first described by Arksey and O'Malley (7) and further advanced by Levac et al. (8) and Peters et al. (9). The methodology of scoping reviews follows a series of steps as follows (7): 1. Identifying the question, 2. Identifying the studies, 3. Selecting studies relevant to the review question from the results of the search, 4. Charting the data, 5. Collating, summarizing, and reporting the findings and 6. An optional consultation with relevant stakeholders. Scoping reviews start with an *a priori* protocol which describes the proposed methodology for each step. A protocol allows for transparency as to which decisions were made *a priori* or during the process of the review itself. Further details on each step of a scoping review are as follows:

1) *Identifying the question*
   The research question for a scoping review is often broad in nature, and is based on the specific objectives of the review. At a minimum, the review question defines the content area and scope of the review. Generally, a scoping review question will define one or two aspects that delineate the scope of the review. Perhaps the easiest approach to understand this is to compare the approach to identifying the review question to the type of question that would be appropriate for a systematic review. Systematic reviews usually are written very precisely to reflect specific key elements of a review question; for intervention questions, these are the population, intervention, comparison, and outcome (see systematic review question types, below, for further detail on key elements). Because a scoping review is describing the literature, rather extracting the study result, a scoping review about an intervention might seek to map this body of literature by defining only the population and the outcome of interest in the scoping review question. For example, while a systematic review, might ask "What the effect of BRD vaccination compared to no vaccination on the incidence of respiratory disease in feedlot cattle," a scoping review might ask, "What interventions have been investigated for the reduction of respiratory disease in feedlot cattle?" In this example, the scoping review has defined the population and outcome, and then will map the literature about the interventions and comparators. Scoping reviews in veterinary medicine have involved a range of species and topic areas, including scoping reviews of the indicators and methods of measurement that have been used to evaluate the impact of population management interventions for dogs (10), non-antibiotic interventions in cattle to mitigate antibiotic resistance of enteric pathogens (11), and indications for acupuncture in companion animals (12).
2) *Identifying the studies*
   The process of searching the literature for relevant studies is the same for scoping and systematic reviews. The intention for a scoping review is to describe the totality of literature on a subject. Thus, the aim is to maximize the sensitivity of the search for identifying relevant literature. Search terms are created to address the key components of the research question, such as the population of interest and the topics

area. These search terms are then combined using Boolean operators and applied to multiple electronic databases as well as other sources such as websites or theses portals (the "gray literature"). The specifics of creating and applying search strategies are consistent with those used in systematic reviews, and so this topic will be more completely covered in later sections of this article.

3) *Selecting relevant studies*
The process of selecting relevant studies is the same for scoping and systematic reviews. Maximizing the sensitivity of the search generally results in a loss of specificity; many non-relevant citations may be captured. Thus, the aim of this step is to identify and remove from the review citations that are not relevant to the scoping review question. This is done by creating a small number (generally one to three) of "screening questions" that can be applied quickly to the titles and abstracts of each citation to allow the identification of citations that are not relevant. The questions often pertain to the population and outcome or topic area of interest. For instance, if the aim of the scoping review is to describe the literature on interventions to prevent respiratory vaccines in swine, the questions might ask whether the citation describes swine as the population of interest, and whether the citation describes the outcome of interest i.e., interventions to prevent respiratory disease. After screening titles and abstracts, full texts are acquired for potentially relevant citations and the screening questions are applied again to the full articles.
To reduce the potential for selection bias in the identification of relevant literature, it is standard practice for relevance screening to be undertaken in duplicate by two reviewers working independently, with any disagreements resolved by consensus. A recent study comparing duplicate screening to limited dual review (only some of the citations screened by two reviewers) reported that up to 9.1% (title and abstract screening) and up to 11.9% (full text screening) of relevant articles were inadvertently excluded when two reviewers were not used (13). However, when the number of citations identified by the search is very large, screening can be undertaken by a single reviewer, with a second reviewer evaluating the studies which were identified as not relevant by the first reviewer. Currently, screening for relevant studies based on the title and abstract is usually conducted by human resources, however machine learning approaches are available to assist in this process, and it is envisioned this process will be fully automated soon.

4) *Charting the data*
This is a step where there are substantial differences between a scoping review and a systematic review. The differences relate to the level of detail extracted and the focus; because they are descriptive, scoping reviews usually do not extract the results of a study and rarely assess the risk of bias in a study (14). For a scoping review, describing the data involves extracting relevant information from each of the articles that have been identified as relevant to the review. The actual information that is collected will depend on the intent of the review as described in the protocol, but often include characteristics of the study (such as location and year), more detailed description of the population (species, stage of production for livestock animals), and the outcomes (potentially including conceptual outcomes, operational outcomes, and outcome measurements such as incidence, prevalence, relative risk or others). Data also may be collected on the aim of each study (e.g., laboratory testing, diagnostic test development, hypothesis testing) and the study design. For example, for a scoping review to address the review question "What interventions have been investigated for the prevention of respiratory disease in swine?", information could be extracted about the population (e.g., stage of production) and possibly further details on the outcome (e.g., identification of specific respiratory pathogens via nasal swaps vs. categorization of lung lesions at slaughter as different operational outcomes for the conceptual outcome of "respiratory disease"), although the broad descriptions of the population and outcomes of interest were already defined in the review question. It is likely that more detail would be extracted related to the interventions and comparators used, because the intent of the review was to explore that aspect of the topic. Data extraction might also include information on the type of study design, if the objective was to identify possible interventions for which there was sufficient data to conduct a systematic review.
Data extraction is usually conducted in duplicate by two independent reviewers using a standardized form developed prior to starting the study, although this form may evolve over the conduct of a scoping review. Disagreements between reviewers are resolved by consensus or with input from a third reviewer.

5) *Collating, summarizing, and reporting the results*
This step also is different from a systematic review, and does not include a meta-analysis. In this step, for a scoping review, the information extracted from each relevant article is collated and presented to the reader. This can be done using tables, figures, and text. The presentation of the information should match the objectives of the scoping study, but may include a description of the type of literature available, changes in the volume or type of literature on the topic over time, or summaries of interventions and outcomes by study design to identify areas where there may be a sufficient body of literature to conduct a systematic review. The PRISMA Extension for Scoping Reviews (PRISMA-ScR) provides guidelines for appropriate reporting of scoping studies (15).

6) *Stakeholder consultation*
The sixth step, which is optional, is to include stakeholder consultation. This may occur at multiple stages of the scoping review (e.g., question formulation, identification of literature, creation of data extraction tools, interpretation of results). As an example, if the scoping review question involved a consideration of management practices at dry-off in cattle, the researchers may consider including a group of dairy veterinarians or producers when discussing the scope of the review, the search terms, and the search strategy. This could help to ensure that all relevant practices are included and that the search terms include both common and potentially less common synonyms for the various management options.

Although this brief summary provides an overview of the steps as they are generally undertaken for scoping reviews, there is a lack of consistency in the terminology and the specific approaches used in studies referred to in the literature as "scoping studies". Colloquially, the process of describing the literature is often called mapping or charting the literature. However, those terms are not well-defined. For example, the American Speech-language-hearing Association seems to equate the term "Evidence Map" with a systematic review (https://www.asha.org/Evidence-Maps/), while the Campbell Collaboration seems to equate the term more closely with a scoping review, describing evidence maps as a "systematic and visual presentations of the availability of rigorous evidence for a particular policy domain" (https://campbellcollaboration.org/evidence-gap-maps.html). There are two published "scoping reviews of scoping reviews" which provide details on how this methodology has been applied in the literature (14, 16) and a discussion of the issue is available by Colquhoun et al. (17). It is likely that the approach to scoping reviews will be further defined and refined over time.

Systematic Reviews

Systematic reviews are intended to summarize the literature to address a specific question. Thus, a systematic review can be seen as an approach to compiling the results from multiple studies addressing the same research question. Detailed descriptions of the methodology as developed for human healthcare questions are available from a number of international consortiums, including the Cochrane Collaboration (18) and the Centre for Reviews and Dissemination (19). A detailed discussion of systematic reviews specific to veterinary medicine is available in a special issue of the journal "Zoonoses and Public Health" (20–25).

As with scoping reviews, the systematic review process follows specific steps. The planned approach for each of the steps is first described in a protocol, which should be completed prior to starting the actual review. Any deviations from the protocol should be acknowledged and justified in the final systematic review report or publication. This transparency allows the reader to understand which decisions were made after the review progress began and reduces the risk of biases, including outcome selection bias (26). Protocols may be published prior to starting the review on websites such as PROSPERO (https://www.crd.york.ac.uk/prospero/) or SYREAF (www.SYREAF.org), posted to University repositories, or submitted as supplementary materials with systematic review publications.

The steps of a systematic review are outlined and briefly discussed, below:

1. ***Define the review question***

   Systematic reviews can be used to address a variety of questions, but not all questions that a veterinarian might wish to have answered can be addressed by a systematic review. Generally, questions where the answer could be expressed as a list are not appropriate for a systematic review (for example: "What vaccines are available for respiratory pathogens in swine?", or "What treatments are used in the management of FUS in cats?"), although these might be appropriate as scoping review questions. Questions that can be addressed with systematic reviews are those that could be answered with a primary research study (18, 27) where the study results estimate a parameter from a sampling distribution. To illustrate using the example of vaccines for respiratory pathogens in swine, the answer to the previously posed question "What vaccines are available?" would be a list of options. However, a related question might be "Does vaccination with *Mycoplasma hyopneumoniae* vaccines at weaning reduce the incidence of respiratory illness during the nursery stage?" For this question, the answer would be an effect size (risk ratio or odds ratio) and an associated measure of variation. Continuing the idea of combining results from multiple studies addressing the same research question, we would expect different studies to provide different estimates from an underlying sampling distribution and the goal often is to summarize the effects, report the average effect size, the observed variation in effect, and factors associated with variation in the average effect size. Some meta-analyses have different underlying assumptions and goals; for a more detailed discussion, see Rice et al. (28).

   In veterinary medicine, systematic review questions generally fall into one of four question types: descriptive questions, intervention questions, exposure or etiology questions, or diagnostic test accuracy evaluations. Each of these question types include "key elements," which should be defined when developing a systematic review question (27).

   i *Descriptive questions.* Systematic reviews may be used to estimate parameters from a single group (effects), such as estimating incidence or prevalence of a condition, or other single group effects such as means or proportions. The key elements that need to be defined for these types of review questions are the population (P) and the outcome(s) (O). Examples of systematic reviews of descriptive questions include estimating the prevalence of *Giardia* in dogs and cats (29), the prevalence of *Campylobacter* in household pets and in petting zoos (30), and the prevalence of *Salmonella* in healthy cattle (31). Sometimes, the outcome of interest may be measured at multiple levels of organization, such as the herd level prevalence and the individual level prevalence of a disease or condition of interest, and possibly a sub-animal unit such as the quarter level in dairy cattle. These technically are different review questions; however, for efficacy, they may be combined into the same workflow process if it is expected that, by and large, the same group of manuscripts will provide the data for both outcomes. For instance, suppose that the review question of interest is "What is the prevalence of intramammary infection with *Staphylococcus aureus* in dairy cattle?" Studies addressing this question might estimate prevalence at the individual quarter level, at the cow level, and at the herd level (perhaps by sampling bulk milk or by defining a cut-point of positive samples necessary to assess a herd as positive). It would not be

sensible to combine results of studies estimating prevalence at these different levels. However, if the results at all levels were of interest to the review team, it might be efficient to conduct a single search of the literature for studies estimating the prevalence of *S. aureus* in dairy cattle without specifying a level in the search terms. In this way, studies estimating prevalence at all levels would be identified, and information on the level of the study could be collected during data extraction and used to conduct separate analyses for each level.

ii *Intervention questions.* A common reason for conducting systematic reviews in veterinary medicine is to synthesize the literature evaluating the efficacy of an intervention. The key elements of this type of question are the population (P), intervention (I), comparison group (C), and outcome(s) (O); thus, review questions for interventions are often referred to by the acronym PICO (or PICOS, if the study design also is identified as a component of the review question). Examples of systematic reviews addressing intervention questions include the efficacy of porcine Circovirus type 2 vaccines in piglets (32), surgical treatments for cranial cruciate ligament disease in dogs (33), and veterinary homeopathy (34). Intervention studies usually report a metric of intervention effect compared across groups such as an odds ratio, hazard ratio, risk ratio, mean difference, or standardized mean difference. As with descriptive questions, some reviews may have multiple outcomes of interest for the same question, and these can be combined into a single review workflow. For instance, a review of the efficacy of vaccines for respiratory disease in calves may include both an outcome related to clinical disease (e.g., the relative risk of treatment with an antibiotic) and a production outcome (e.g., the mean difference in average daily gain between treatment groups).

iii *Exposure questions.* Systematic reviews also can be used to address questions related to etiology or exposures (including dose-response), with the key elements for these types of questions being the population (P), exposure (E), comparison (C), and outcome(s) (O). Examples of exposure review questions include risk factors for *Salmonella* in laying hens (35) and risk factors associated with transmission of *Mycobacterium avium subsp. paratuberculosis* to dairy calves (36). Exposure studies usually report a metric of intervention effect compared across groups such as an odds ratio, hazard ratio, risk ratio, mean difference, or standardized mean difference.

iv *Diagnostic test accuracy questions*. Systematic reviews may be used to synthesis the available literature to determine diagnostic test accuracy. The key elements of this type of review question are the population (P), index test (I), and target condition (T). Recent examples of diagnostic test accuracy systematic review include reviews to estimate the diagnostic accuracy for detecting bovine respiratory disease in feedlot cattle (37) and to compare diagnostic tests for reproductive tract infections and inflammation in dairy cows (38). Diagnostic test accuracy review questions often report metrics of test performance such as sensitivity and specificity or likelihood ratios.

TABLE 1 | Example of a simple search strategy that could be used to identify studies evaluating the efficacy of probiotics to reduce or prevent diarrhea in horses using Medline via PubMed.

Search	Query	Items found
#1	Horse or horses or pony or ponies or donkey or donkey or equine	98,264
#2	Probiotic or probiotics or yeast or lactobacillus or "lactic acid bacteria" or bifidobacteria or Saccharomyces or "*Bacillus subtilis*"	367,665
#3	Diarrhea or enteric or gastrointestinal or GI or scours	532,459
#4	#1 and #2 and #3	60

2. ***Conduct a comprehensive search for studies***

As with literature searches for scoping reviews, the aim is to identify all of the available literature on the topic. Once the systematic review question has been defined, a list of search terms is created using some or all of the PICO elements (for intervention questions), including their synonyms or other related words. The words are then combined using Boolean operators such as "AND," "OR," or "NOT" to create search strings (39). Filters can be applied to limit the search by year of publication, language of publication, publication type, or study design. Search strings are then applied using a search strategy to identify potentially relevant studies. Searching the veterinary literature can be challenging, in that search filters used in human healthcare literature searches are lacking for veterinary medicine and reporting of key features in titles and abstracts may be poor (40).

A simple example of a search string as applied in Medline via Pubmed is provided for the review question "What is the efficacy of probiotics compared to no treatment for reducing or preventing diarrhea in horses" (**Table 1**). Key words are included for the population (horse), intervention (probiotics) and the outcome (diarrhea). Within each of these key element concepts, search terms are linked with "OR," meaning that eligible citations need only include one of these words to be identified by the search. The key element concepts are linked with "AND," meaning that the citations need to include at least one word in each of the key concept blocks. In this example, we have included plural forms and words; however, different databases will have symbols which allow truncation of words, as well as other features to enhance the search process. Additionally, this example is quite simplistic; key words may have been missed, and the actual syntax of the search string will differ between databases. Because of the complexity of literature searching, and the importance of identifying all of the relevant literature, including a library scientist on the review team to assist with the development of the search string and the application of the search strategy can be extremely helpful.

Searches should be designed to capture both journal articles indexed in electronic databases and other types of research reports such as theses, government reports, and conference

proceedings (referred to as the "gray literature"). It is recommended that the search include multiple databases. The electronic databases that are appropriate for systematic reviews in veterinary medicine may differ from those in human healthcare. Grindlay et al. (41) evaluated available electronic databases for coverage of veterinary journals and found that the highest journal coverage was for the Scopus database and the Cambridge Agricultural and Biological Abstracts Index. Searching the gray literature can be challenging, and the assistance of a library scientist is recommended. A recent discussion of gray literature searching, including a list of resources for searching the gray literature, is provided by Paez (42). The usefulness of available resources for searching the gray literature specifically to identify veterinary research has not been evaluated. Another approach (not mutually exclusive) is to search the reference list of recent review articles or of articles identified as relevant to the review.

Once the searches have been conducted, the citations identified by the search are uploaded into a reference management software. Because of overlap in journal coverage between databases, there will likely be duplication citations identified by the search. Most reference management software will have an internal program that can be used to identify and remove duplicate records.

3. ***Select relevant studies from the search***

Systematic literature searches, as with scoping review searches, are designed to maximize the sensitivity in identifying the literature of relevance to the review questions. However, this means that the specificity is often low. Thus, citations must be screened to ensure that they meet eligibility criteria for the review. The process of eligibility (or "relevance") screening follows the same procedures as was described previously for a scoping review.

4. ***Collect data from relevant studies***

Information of relevance to the review question is extracted from each study. This includes information at the study level such as year of publication, months, and years when the study was undertaken, study design (if multiple study designs are eligible), and geographic region. Information on the study population and, for PICO questions, the intervention and comparison groups may be collected for two reasons; they provide the necessary context for the reader to interpret the results, and they may be evaluated as possible explanations for any differences identified among the included studies. Determining *a priori* what information on the population and the intervention groups is of interest can be challenging, and it is helpful to have someone with content expertise on the review team to assist with these decisions. As with all aspects of a systematic review, these decisions should be made during the protocol development stage.

Information related to the results of the study (or, for intervention studies, each comparison) also needs to be extracted. This includes the characteristics of the outcomes (e.g., the case definition, the method of assessing the outcome) and the result; generally, the sample size and effect in each group or the effect size, sample size, and a measure of variability. It is common for studies in the veterinary sciences to report a large number of outcomes; in clinical trials in small companion animal populations, the mean number of outcomes per trial was 10.8 and ranged from 1 to 30 (43), and in clinical trials in livestock, the mean number and range of outcomes was 8.5 and 1–41, respectively (44). Therefore, during the protocol development stage, the investigators need to decide which outcomes to use, and how many to include. Although there is no set rule, the selected outcomes should be those of relevance to decision-making and could pertain to possible harms as well as benefits. Including too few outcomes may not provide the decision-maker with enough information to make an evidence-based decision; but including too many outcomes is time intensive and may lead to a lack of focus in the review. The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) working groups recommends including up to seven critical outcomes in their summary of findings tables for interpreting the quality of evidence from a systematic review (45).

To reduce the potential for misclassification, data extraction should be conducted by two reviewers working independently with disagreements resolved by consensus, or by a single reviewer with a second reviewer validating the information extracted by comparing to the full text report, or a single reviewer after a period of duplicate extraction with verification of consistency.

5. ***Assess risk of bias in relevant studies***

Assessing the risk of bias in the primary studies that are included in a systematic review allows for a consideration of bias in the interpretation of the results, and is a step not normally included in scoping reviews. At this stage of the review, it is not appropriate to remove studies from a review based on the risk of bias, unless one or more trial design features are included *a priori* in the eligibility criteria (for instance, eligible study designs were defined as trials with random allocation to treatment groups). However, understanding the potential for bias helps to interpret the quality of the evidence produced by a review, and factors related to risk of bias may be explored in a meta-regression or sub-group meta-analysis as possible reasons for differences in results among included trials.

The actual criteria used in assessing bias will differ by study design. Validated tools for assessing risk of bias are available for different study designs (see: https://www.riskofbias.info/). A commonly used instrument for assessing the risk of bias in clinical trials is the Cochrane Risk-of-Bias 2.0 tool (46). In this tool, the risk of bias is assessed in five domains using signaling questions to assist the reviewer in determining a judgement for each domain. The signaling questions within each domain are answered as yes, probably yes, no, probably no, or no information. These signaling questions have been designed to make the judgement about risk of bias more consistent and reproducible (46). For each domain, an algorithm is then applied using the answers to the signaling question to determine whether the risk of bias for that domain is high, some concerns, or low. The domains relate to the risk of bias arising from the randomization process, the risk of bias

due to deviations from the intended interventions, the risk of bias due to missing outcome data, the risk of bias due to the measurement of the outcome, and the risk of bias in the selection of the reported result. An overall risk of bias for the trial can then be assessed as low, some concerns or high. Some modifications to the tool may be necessary for evaluating trials conducted in some livestock populations. For instance, when assessing the risk of bias in trials in swine populations, Moura et al. (47) did not include allocation concealment in their algorithm, because the authors felt that this was unlikely to be an essential design feature for populations where all eligible pens were allocated to groups, with no reason for any *a priori* preference as to treatment group.

Other tools are available for assessing the risk of bias for observational studies. The ROBINS-I tool was developed for assessing the risk of bias in non-randomized studies of interventions (48). Robins-I is designed for exposures that could be randomly allocated "based on a hypothetical pragmatic randomized trial." Such a "target" trial need not be feasible or ethical: for example, it could compare individuals who were and were not assigned to start smoking. For some exposures, such as sex, age, region, or production stage, this hypothetical trial concept is not applicable; therefore, Robins-I is easiest to use when the intervention could actually theoretically be randomly allocated. Robins-I also uses signaling questions to aid the reviewer in the assessment of risk of bias, but across seven domains of relevance to non-randomized studies (48). The domains of bias assessed at the pre-intervention stage are bias due to confounding and bias due to selection of participants into the study, at the intervention stage the potential for bias due to classification of the intervention is assessed, and at the post-intervention stages the domains assessed are bias due to deviations from intended interventions, bias due to missing data, bias in measurement of outcomes, and bias in the selection of the reported results. As with the Cochrane Risk-of-Bias 2.0 tool, an overall risk of bias determination is made based on the risk of bias across the seven domains. Other tools are available for assessing the risk of bias in observational studies, such as the Newcastle-Ottawa quality assessment scale (49) and the RTI item bank (50). For systematic reviews of descriptive questions, a modification of the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool has been proposed for assessing the risk of bias of primary descriptive studies (51). For exposure that cannot be randomized, risk of bias tools still require validation.

6. ***Synthesize the results***

Systematic reviews may include a qualitative synthesis or a quantitative synthesis of the results of the primary studies that were included in the review. If the studies are too disparate to justify combining then to a common result, it is still of value to present the results of the eligible studies qualitatively using tables and text. Meta-analysis provides a weighted average of the results of the individual studies (18). For intervention (PICO) questions, the results of a study are a comparison of two groups. These comparative measures are often referred to in a non-specific manner as effect sizes, a terminology that arises from clinical trials where, due to random allocation, the difference in groups is interfered as the effect of the intervention. The term effect size refers to any measure used to compare two groups. Common measures are the odds ratio, risk ratio, mean differences, and correlation. It is also possible to conduct a meta-analysis on diagnostic test evaluations. The measures used in these studies differ from group comparisons and include sensitivity, specificity, correlation, and the ROC curve. The results of several descriptive studies can also be combined as a weighted average. Common measures summarized across multiple descriptive studies include prevalence or incidence or dose response.

The common components of a meta-analysis are the calculation of a summary effect or effect size, an evaluation of the heterogeneity of results (differences in the effect or effect size among studies), a visual presentation of these results using a forest plot, and an evaluation of the potential for small study effects ("publication bias"). Each of these is described briefly below using hypothetical data examining the odds ratio for "clinical cure" (measured as a binary variable) associated with a new intervention compared to the existing standard of care. An example dataset and R-code to calculate the various components of the example meta-analysis is included as Appendix 1 in the **Supplementary Material**. The interested reader can find additional details on the methodology underlying meta-analysis in Higgins and Green (18), CRD (19), and O'Connor et al. (25). We present the results using odds ratios as the outcome measure, which were calculated from arm-level data. Sometimes data are present in a study as a comparative measure, such as unadjusted or adjusted odds ratios, as opposed to arm level data. If the meta-analysis contains studies with both arm level and adjusted odds ratio, it will be necessary to convert the arm level data to the odds ratio scale because it is not possible to convert an adjusted odds ratio back to arm level data. For a meta-analysis, the data from all included studies need to be in the same form i.e., either all arm level data or all odds ratios with a measure of variability.

Meta-analysis of binary data from two groups is usually conducted on the log odds scale (i.e., the difference in the log odds). The software package recognizes if the data are arm level or contrast level and conducts this conversion. After the analysis is complete, the data are usually converted back to the odds ratio and sometimes back to the risk ratio using the expit formula (inverse of the logit function). The presentation of risk ratio results based on back transformation of log(odds) is quite different from direct meta-analysis of the log(risk ratio) which is less common because the risk ratio has mathematical constraints that can create bias in the meta-analysis [see Bakbergenuly et al. (52)] for a discussion of this topic. Although meta-analysis of binary data from two groups is usually conducted using the odds ratio as the metric, the issue of non-collapsibility remains in meta-analysis as it does with primary research [see Rothman et al. (53) for a discussion of this topic in the primary research and Bakbergenuly et al. (52) for a discussion of this topic in meta-analysis].

The first step of a meta-analysis is to calculate a weighted average of the effect size and then convert that, if necessary, back to a scale of interest. In our example, the meta-analysis calculates the average weighted log odds ratio and then converts that to a summary odds ratio. A meta-analysis can either calculate the weighted mean effect size using a fixed effect(s) approach [where it is assumed that the true effect size is a single common value or several single effects (28)] or using a random effects approach (where it is assumed that the true effect size follows a distribution). A random effects approach is usually considered most appropriate in the situation where the observed studies are considered to be a representative sample of the population. The fixed effect(s) approach is appropriate if the goal is to make inference conditional on the observed studies (54). Further, if the difference in effects observed in studies are not regarded as random, then a fixed effect(s) approach to analysis may be suitable. Rice et al. (28) also discusses the two underlying data generating mechanisms that can be used to make inference from a fixed effect(s) approach. The first is that there is a single true effect and sampling error causes observed differences in the estimates. This approach is discussed in detail by Borenstein et al. (55), and is referred to as the fixed effect (note singular) approach. However, Rice et al. (28) have recently proposed referring to this as a common effect model (54). The second hypothesized data generating mechanism under the fixed effect(s) approach is that the studies are estimating different effects, i.e., the variation is due to different effects and is not random. This is referred to by Rice et al. (28) as a fixed effects (note plural) approach. There has been debate over the relative merits of fixed(s) and random-effects approaches to meta-analysis and there is no consensus as to which approach is appropriate and under what circumstances (28).

With a fixed effect approach, studies are frequently weighted based on the inverse of their variance; thus, larger studies tend to contribute more to the summary effect size than smaller studies (55). Another approach to fixed effects meta-analysis is the Mantel-Haenszel, which is used when data are sparse and requires a different weighting approach based on the summary statistics and in particular uses a weighted odds ratio rather than a log odds. The Peto method of fixed effect meta-analysis, also called the one-step method, uses the log odds ratio and a variant of the inverse variance weighting approach (55). For the random effects approach, both the within study variance and the between study variance are considered in the weighting. The method of DerSimonian and Laird (56) is a commonly used method to estimate the between and within study variance because it is the default approach in many software packages. However, there are alternative approaches to estimation that are may be better and are available in many packages (57). In the meta-analysis section of the Cochrane Handbook (58), it has been proposed that the approaches by Hartung and Knapp (59) and Sidik and Jonkman (60) should be used if available to review authors as the confidence intervals are correctly adjusted to account for the uncertainty associated with the estimation of the between study variance.

The results of a meta-analysis often are displayed using a forest plot (**Figure 1**). In this plot, the results of each study are shown both numerically (columns on the right of the figure) and graphically. In the figure, the results of each comparison are shown by a box (representing the point estimate of the effect size) and by a horizontal line (representing the 95% confidence intervals). On the right of the graph, the weighting of each comparison to the final summary effect size is shown. At the bottom of the plot, the summary effect size is shown numerically (point estimate and 95% confidence intervals) and graphically, using a diamond, with the center of the diamond representing the point estimate and the horizontal ends of the diamond representing the 95% confidence intervals on the estimated mean summary effect size. In the example dataset, and using a random effects approach, the summary odds ratio was 1.58 (95% confidence intervals: 1.01, 2.48). As with primary research, the confidence interval relates to uncertainty around the average effect of the distribution and does not describe the variation in the underlying distribution of the "random effect." Tau squared is the between study variance and the square root of tau squared, tau, is the estimate of the standard deviation across the studies. If it is of interest to describe to the reader the distribution of the studies (i.e., how much study effects vary) this can be reported by using a prediction interval. The prediction interval incorporates two levels of variation, the standard error of the estimated weighted mean of the distribution and the estimate of between study variance Tau squared (61).

Visually assessing the forest plot in **Figure 1**, it is apparent that not all of the individual studies observed the same results; in some of the studies, it appeared that the new treatment was better and in some worse, and for some of the studies the 95% confidence intervals included the null value (odds ratio of 1) whereas in other studies it did not. There are two common measures used to quantify heterogeneity in the results of a meta-analysis. The first of these is Cochran's Q statistic and corresponding Chi-square based test which evaluates the homogeneity of the effect size of the studies (62). While informative, the Cochrans Q test tends to be of low power, as the number of studies included in most meta-analyses is quite low. The second measure is I^2 which tells us the relationship between the two sources of variation that we expect in a meta-analysis—the variation of true effects and the variation due to sampling error (62). As such, I^2 describes the percentage of total variation across studies that is due to heterogeneity in effects rather than chance. I^2 is frequently reported as describing how much the effects vary. However, I^2 is a proportion rather than an absolute value. If the I^2 percentage is small, this implies that if all the sampling error was removed, the true effects would not differ greatly i.e., consistent true effects. If the I^2 percentage is large, this implies that if all the sampling error was removed, the true effects would differ greatly among studies i.e., more variation due to true effects (62). Guidelines are available to interpret I^2; values of 0–40% are likely unimportant, 30–60% represents moderate heterogeneity, 50–90% represents substantial heterogeneity, and 75–100% represents considerable heterogeneity (18).

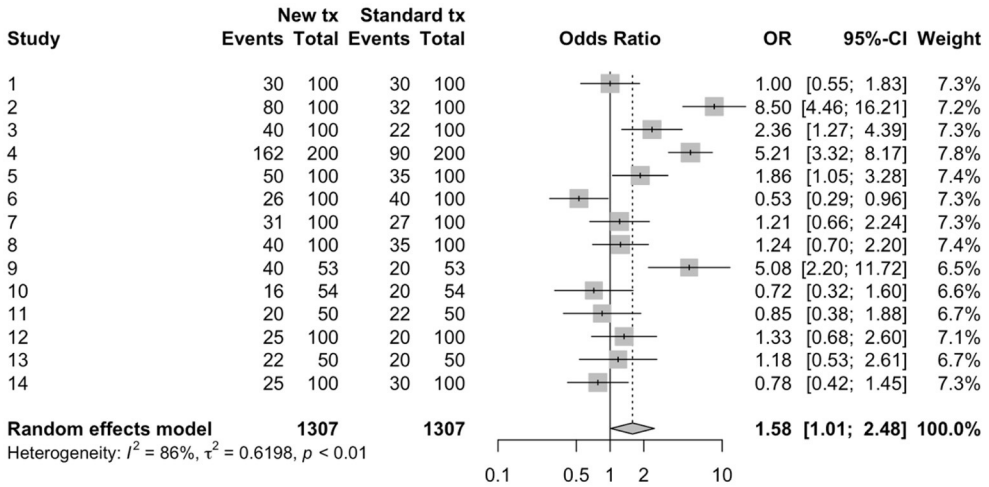


FIGURE 1 | Forest plot of a meta-analysis conducted using arm-level data to estimate odds ratios from 14 hypothetical trials comparing a new treatment to a standard treatment.

In the example meta-analysis, the I^2 shows that 86% of the variability in the individual study results were due to heterogeneity, rather than chance. Thus, quantifying the summary effect size might not be useful for this example, and exploration of factors that contribute to the heterogeneity might enable better understand the effect size and factors impacting the effect size.

Beyond sampling error, heterogeneity may be related to clinical (contextual) or methodological factors. Clinical heterogeneity results from variability in the population, intervention, or outcome, whereas methodological heterogeneity results from variability in study design or risk of bias among studies (18). Techniques are available to explore possible sources of heterogeneity. Subgroup meta-analysis can be used to explore suspected sources of heterogeneity by dividing the primary studies into subgroups based on the characteristic that is thought to be a source of heterogeneity (18). These potential sources of heterogeneity may be clinical or methodological. In the working example, a subgroup meta-analysis was conducted based on whether or not the study employed random allocation of study subjects to treatment group (**Figure 2**). In this figure, we can view the meta-analysis (and resulting evaluations of heterogeneity) separately for trials that used random allocation to treatment group (the lower forest plot) and those that did not (the upper forest plot). In this hypothetical example, we see that the results differ; the meta-analysis of non-randomized trials showed a large beneficial effect of the new treatment vs. the standard treatment (summary OR = 4.79, 95% CI = 2.87, 7.99), although the I^2 value was still high at 63%. However, for the randomized studies, the point estimate of the summary OR suggests no benefit (1.04) and the 95% confidence interval is quite precise (0.81–1.30) suggesting that there is no evidence, on average, of benefit to the new treatment. In the meta-analyses of randomized studies, the I^2 value indicates that heterogeneity was not a concern.

Another approach to exploring heterogeneity is meta-regression. Meta-regression is a weighted regression of the results of the individual studies (the unit of concern) on the variables of interest, which often are possible sources of heterogeneity. As with meta-analysis, the weighting generally is the inverse variance of each study's result. Details on how to conduct univariable or multivariable meta-regression can be found elsewhere (18, 25).

Publication bias is a potential concern whenever research results are considered in decision-making. Publication bias occurs because studies showing preferred results are more likely to be published or are published faster (63). We use the term "preferred result" rather than positive or negative result, as these terms can be misleading and have multiple meanings. For example, for an outcome such as mean difference, the results can be positive or negative, but depending upon the outcome, a negative or positive mean difference might be preferred. Another use of the term positive or negative might be to confer inference; for example, a vaccine a product intended to present a specific disease that has an OR greater than one compared to an untreated control has a "positive" result in absolute terms. If the outcome was mortality, a "positive" OR would not be the preferred outcome. However, if the outcome was the probability of sero-converting to the disease agent of interest, then an OR of greater than one would be the preferred outcome.

There is empirical evidence that many research studies in the veterinary sciences are not published in the peer-review literature; a study of conference proceedings abstracts for swine and cattle vaccine trials found that <10% of the studies were subsequently published (64), and interventions related to on-farm and abattoir food safety reported that less than half of the research was published in the peer-reviewed literature

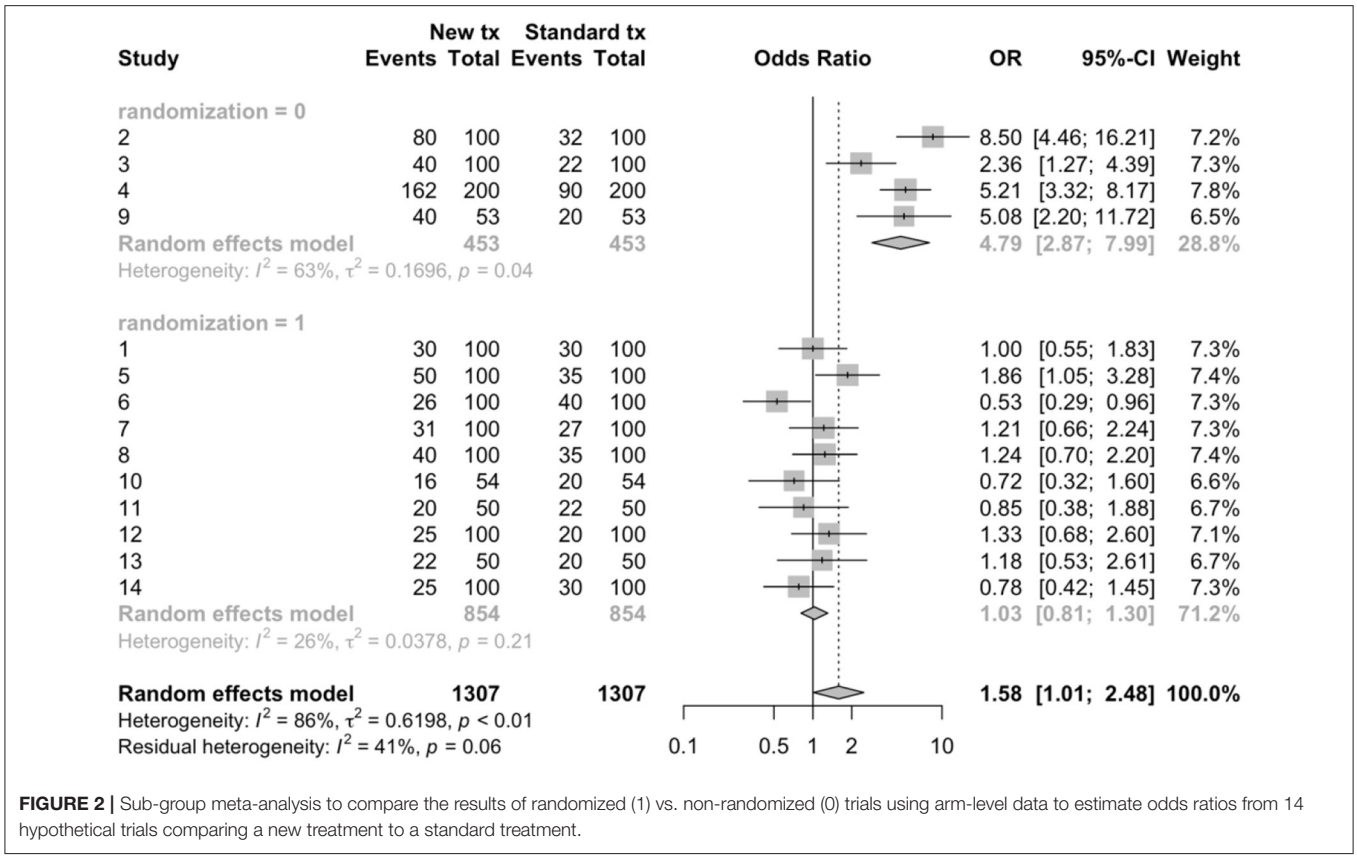


FIGURE 2 | Sub-group meta-analysis to compare the results of randomized (1) vs. non-randomized (0) trials using arm-level data to estimate odds ratios from 14 hypothetical trials comparing a new treatment to a standard treatment.

within 4 years (65). Small study effects, one explanation for which is publication bias, can be assessed using a funnel plot, which plots the effect size from individual studies on the x-axis and some measure of variability (inverse variance, standard error, inverse of the standard error, or sample size) on the y-axis (63, 66, 67). The resulting figure is called a funnel plot because the precision of an effect size increases as the sample size increases. Thus, it would be expected that smaller studies would have a wider range of estimates and larger studies a smaller range of estimates, leading to a funnel shaped plot in the absence of publication bias. Publication bias will result in an asymmetric shape, with smaller non-significant publications not represented, although it should be noted that there are reasons other than publication bias that may result in a non-symmetric funnel (67). An example might be if both challenge studies and natural disease exposure trials are included in the same review; challenge studies tend to be smaller and also tend to report a larger effect size (68). This is an example of a source of heterogeneity whereby small studies may have a different effect compared to larger studies. Funnel plots represent a visual approach to detecting publication bias, although it may be difficult to accurately assess whether publication bias is present or not based on a visual appraisal (69). There are statistical tests available to formally evaluate asymmetry in a funnel plot, the most common being a rank correlation test and the regression test (63, 70, 71). Other approaches to detecting publication bias include the selection model approach. Selection models use the weighted distribution theory to model the selection and can be complicated, especially compared to the funnel plot approach. A comprehensive review of this approach is available (72). In addition to detecting publication bias, it might be interest to quantify the effect of publication bias on the effect size and adjust for the bias. A review of these methods is available elsewhere (73). A funnel plot for the example data using standard error as the measure of variability on the y-axis is shown in **Figure 3**.

7. ***Present the results***

The results of a meta-analysis generally are presented using text, figures, and tables. In most cases, the presentation of results will include a table summarizing the study characteristics, as well as tables and figures showing the individual study results and the results of risk of bias assessments. If a formal meta-analysis was conducted, forest plots and funnel plots also may be shown. A summary of the findings of a systematic review by outcome may be presented in a "summary of findings" table (45). These tables include the summary effect size with 95% confidence intervals, the total number of participants and studies, an estimate of absolute effect, and the overall quality of evidence (see below).

The methods and results of a systematic reviews should be reported in sufficient detail that the reader can evaluate the potential for bias. The Preferred Reporting of Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines

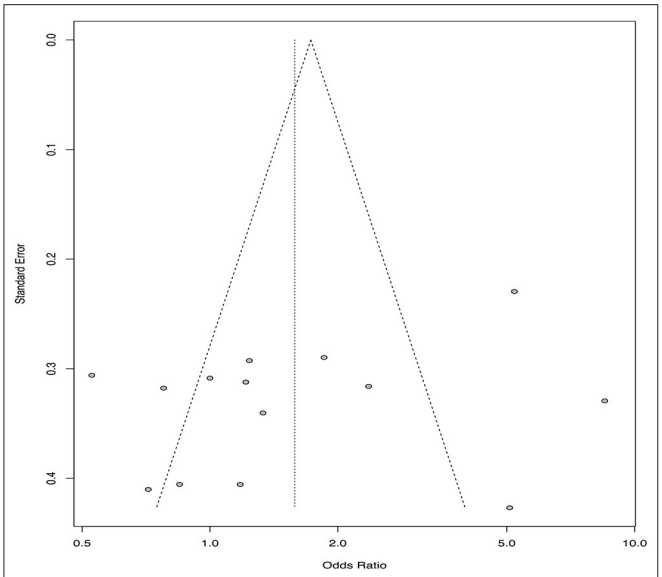


FIGURE 3 | Funnel plot from 14 hypothetical trials comparing a new treatment to a standard treatment using standard error on the y-axis and odds ratio estimates from each included trial on the x-axis.

provide recommendations for the items of information and the level of detail that should be included in a systematic review report [(74, 75); www.prisma-statement.org].

8. ***Interpret the results***

If one or more meta-analyses are conducted, the results should be interpreted in the context of the magnitude of effect and the confidence in the evidence (quality of evidence). The magnitude of effect relates to the summary effect size and its variability. One approach to evaluating the quality of evidence is by using the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) (76, 77). The overall quality of evidence is determined to be high, medium, low, or very low. The framework evaluates the quality of evidence by considering four domains; risk of bias, publication bias, imprecision, inconsistency, and indirectness. If there are serious concerns in a domain, the overall quality of evidence will be downgraded by one level; if the concerns are very serious within a domain, the quality of evidence can be downgraded by two levels. Reviews of randomized trials start at the rating of "high," whereas observational studies start at a level of "low." Observational studies may be upgraded, as well as downgraded. The domains for evaluating quality of evidence, with references to further details for each, are as follows:

a) *Risk of bias* (78). In this domain, the risk of bias in the individual studies is considered. If most of the evidence is from studies with a high risk of bias, the reviewer may wish to downgrade the quality of evidence. For instance, if the majority of evidence in a review came from trials where allocation to treatment group was not random, the quality of evidence would be lower than a review where the individual trials employed random allocation to treatment group.

b) *Publication bias* (79). If there is strong evidence of publication bias, to the point where the reviewer believes that it may have impacted the review results, the quality of evidence may be downgraded.

c) *Imprecision* (80). Imprecision may be a concern if the overall sample size is less that the sample size that would be appropriate to address the research question in a single study, or if the confidence intervals on the summary effect size span harm, no association, and benefit.

d) *Inconsistency* (81). Inconsistency is related to the heterogeneity in the results. If there was considerable heterogeneity in the results, but the reviewers were able to explain the sources of that heterogeneity (for instance, using subgroup meta-analysis), then inconsistency may not be a concern. However, if there is substantial unexplained heterogeneity, then the quality of evidence may be downgraded.

e) *Indirectness* (82). Indirectness relates to the applicability of the evidence to the research question. Indirectness may relate to one or more of the PICO elements. For instance, if the question of interest was the efficacy of a treatment in market weight pigs, and yet most of the studies identified evaluated the treatment at the start of the finishing period, the evidence would be less direct.

Examples of the use of GRADE in veterinary systematic reviews include a review of furosemide for exercise-induced pulmonary hemorrhage in racehorses (83), a review of the of the efficacy of whole-cell killed *Tritrichomonas foetus* vaccines in beef cattle (84), and a review of on-farm interventions to reduce *Salmonella* in swine (85).

FUTURE DIRECTIONS

There are a number of exciting enhancements to systematic reviews that are recently available or are likely to be available in the new future. These include automation of some of the steps of a systematic review, "living systematic reviews" which evolve as new information becomes available, and network meta-analysis, where multiple intervention options for the same outcome can be assessed in a single analysis which includes both direct and indirect evidence.

Systematic reviews not only are time sensitive, but also require a considerable input of time and resources, taking an average of 67 weeks from protocol registration to publication (86). Thus, having mechanisms to semi-automate at least some stages of a review using machine learning or natural language processing is inherently appealing. Recently, the journal "Systematic Reviews" published an editorial and three commentaries outlining the state of automation in systematic reviews (87–90). Existing tools for using automation in systematic reviews are available and may be used to identify randomized controlled trials, or for eligibility screening, data extraction, or risk of bias. However, there is a need to validate existing tools and continue to develop the techniques. This is a rapidly developing area in systematic reviews and it is

likely that automation of some steps of a systematic review will become more common over time.

Systematic reviews provide a rigorous method of synthesizing the literature, but they represent the evidence at a static point in time. Research is identified up until the date that the literature search was conducted. However, it then takes time to conduct the review and for the review to make its way through the publication process. In human healthcare, it has been estimated that reviews are seldom updated within 2 years of publication (91). Thus, the information available in published systematic reviews may not represent the current state of knowledge (92). Nonetheless, updating systematic reviews requires time and resources, and periodic updating does not negate the time between the conduct of a literature search and formal publication (92). However, technological advances, including automation, allow for the creation of living systematic reviews (92). Unlike traditional systematic reviews, which are in the form of written reports or publications, living systematic reviews are dynamic on-line evidence summaries which can be updated frequently and rapidly as new information becomes available in the literature (92). As the technology and approach become more common, living systematic reviews have the potential to provide the most up to date and rigorous summary of the evidence possible to assist veterinarians and producers in making the most evidence-informed clinical decisions possible.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

JS and AO'C co-determined the content and scope of the review. JS drafted the manuscript. AO'C reviewed the manuscript. All authors approved the final contents.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.00011/full#supplementary-material

REFERENCES

1. Masic I, Miokovic M, Muhamedagic B. Evidence based medicine - new approaches and challenges. *Acta Inform Med.* (2008) 16:219–25. doi: 10.5455/aim.2008.16.219-225
2. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. (1996) 13:312:71–2. doi: 10.1136/bmj.312.7023.71
3. Ioannidis JPA. Evolution and translation of research findings: from bench to where? *PLoS Clin Trials*. (2006) 1:e36. doi: 10.1371/journal.pctr.0010036
4. Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. *Lancet*. (1998) 351:123–7. doi: 10.1016/S0140-6736(97)08468-7
5. Gehr BT, Weiss C, Porzsolt F. The fading of reported effectiveness. A meta-analysis of randomised controlled trials. *BMC Med Res Methodol*. (2006) 6:25. doi: 10.1186/1471-2288-6-25
6. Athanasiou T, Darzi A. *Evidence Synthesis in Healthcare: A Practical Handbook for Clinicians*. New York, NY: Springer Science & Business Media (2011).
7. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int. J. Soc. Res. Methodol.* (2005) 8:19–32. doi: 10.1080/1364557032000119616
8. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci.* (2010) 5:69. doi: 10.1186/1748-5908-5-69
9. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. *Int J Evid Based Health*. (2015) 13:141–6. doi: 10.1097/XEB.0000000000000050
10. Hiby E, Nattrass Atema K, Brimley R, Hammond-Seaman A, Jones M, Rowan A, et al. Scoping review of indicators and methods of measurement used to evaluate the impact of dog population management interventions. *BMC Vet Res*. (2017) 13:143. doi: 10.1186/s12917-017-1051-2
11. Murphy C, Fajt V, Scott H, Foster M, Wickwire P, McEwen S. Scoping review to identify potential non-antimicrobial interventions to mitigate antimicrobial resistance in commensal enteric bacteria in North American cattle production systems. *Epidemiol Infect*. (2016) 144:1–18. doi: 10.1017/S0950268815000722
12. Rose WJ, Sargeant JM, Hanna WJB, Kelton D, Wolfe DM, Wisener LV. A scoping review of the evidence for efficacy of acupuncture in companion animals. *Anim Health Res Rev*. (2017) 18:177–85. doi: 10.1017/S1466252317000068
13. Stoll CRT, Izadi S, Fowler S, Green P, Suls J, Colditz GA. The value of a second reviewer for study selection in systematic reviews. *Res Synth Methods*. (2019) 10:539–545. doi: 10.1002/jrsm.1369
14. Pham MT, Rajic A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. *Res Synth Methods*. (2014) 5:371–85. doi: 10.1002/jrsm.1123
15. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. (2018) 169:467–73. doi: 10.7326/M18-0850
16. Tricco AC, Lillie E, Zarin W, O'Brien K, Colquhoun H, Kastner M, et al. A scoping review on the conduct and reporting of scoping reviews. *BMC Med Res Methodol*. (2016) 16:15. doi: 10.1186/s12874-016-0116-4
17. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. *J Clin Epidemiol*. (2014) 67:1291–4. doi: 10.1016/j.jclinepi.2014.03.013
18. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. The Cochrane Collaboration (2011). Available online at: http://handbook.cochrane.org (accessed October 3, 2019).
19. Centre for Reviews and Dissemination, University of York. *CRD's Guidance for Undertaking Reviews in Health Care*. (2009). Available online at: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf (accessed October 3, 2019).
20. Sargeant JM, O'Connor AM. Introduction to systematic reviews in animal agriculture and veterinary medicine. *Zoon Public Health*. (2014) 61:3–9. doi: 10.1111/zph.12128
21. Sargeant JM, Kelton DF, O'Connor AM. Study designs and systematic review of interventions: building evidence across study designs. *Zoon Public Health*. (2014) 61:10–7. doi: 10.1111/zph.12127
22. Sargeant JM, Kelton DF, O'Connor AM. Randomized controlled trials and challenge trials: design and criterion for validity. *Zoon Public Health*. (2014) 61:18–27. doi: 10.1111/zph.12126
23. O'Connor AM, Anderson KM, Goodell CK, Sargeant JM. Conducting systematic reviews of intervention questions I: writing the review protocol, formulating the question and searching the literature. *Zoon Public Health*. (2014) 61:28–38. doi: 10.1111/zph.12125
24. Sargeant JM, O'Connor AM. Conducting systematic reviews of intervention questions II: relevance screening, data extraction, assessing the risk of bias,

presenting the results and interpreting the findings. *Zoon Public Health*. (2014) 61:39–51. doi: 10.1111/zph.12124
25. O'Connor AM, Sargeant JM, Wang C. Conducting systematic reviews of intervention questions III: synthesizing data from intervention studies using quantitative approaches (meta-analysis). *Zoon Public Health*. (2014) 61:52–63. doi: 10.1111/zph.12123
26. The PLoS Medicine Editors. Best practice in systematic reviews: the importance of protocols and registration. *PLoS Med*. (2011) 8:e1001009. doi: 10.1371/journal.pmed.1001009
27. European Food Safety Authority. Application of systematic review methodology to food and feed safety assessments to support decision making. *ESAF J*. (2010) 8:1637–727. doi: 10.2903/j.efsa.2010.1637
28. Rice K, Higgins JPT, Lumley T. A re-evaluation of fixed effect(s) meta-analysis. *J R Stat Soc Ser A Stat Soc*. (2018) 181:205–27. doi: 10.1111/rssa.12275
29. Bouzid M, Halai K, Jeffreys D, Hunter PR. The prevalence of Giardia infection in dogs and cats, a systematic review and meta-analysis of prevalence studies from stool samples. *Vet Parasitol*. (2015) 207:181–202. doi: 10.1016/j.vetpar.2014.12.011
30. Pintar KD, Christidis T, Thomas MK, Anderson M, Nesbitt A, Keithlin J, et al. A systematic review and meta-analysis of the *Campylobacter* spp. prevalence and concentration in household pets and petting zoo animals for use in exposure assessments. *PLoS ONE*. (2015) 10:e0144976. doi: 10.1371/journal.pone.0144976
31. Gutema FD, Agga GE, Abdi RD, De Zutter L, Duchateau L, Gabriël S. Prevalence and serotype diversity of salmonella in apparently healthy cattle: systematic review and meta-analysis of published studies, 2000-2017. *Front Vet Sci*. (2019) 6:102. doi: 10.3389/fvets.2019.00184
32. da Silva N, Carriquiry A, O'Neill K, Opriessnig T, O'Connor AM. Mixed treatment comparison meta-analysis of porcine circovirus type 2 (PCV2) vaccines used in piglets. *Prev Vet Med*. (2014) 117:413–24. doi: 10.1016/j.prevetmed.2014.10.006
33. Bergh MS, Sullivan C, Ferrell CL, Troy J, Budsberg SC. Systematic review of surgical treatments for cranial cruciate ligament disease in dogs. *J Am Anim Hosp Assoc*. (2014) 50:315–21. doi: 10.5326/JAAHA-MS-6356
34. Mathie RT, Clausen J. Veterinary homeopathy: systematic review of medical conditions studied by randomised trials controlled by other than placebo. *BMC Vet Res*. (2015) 11:236. doi: 10.1186/s12917-015-0542-2
35. Denagamage T, Jayarao B, Patterson P, Wallner-Pendleton E, Kariyawasam S. Risk factors associated with *Salmonella* in laying hen farms: systematic review of observational studies. *Avian Dis*. (2015) 59:291–302. doi: 10.1637/10997-120214-Reg
36. Doré E, Paré J, Côté G, Buczinski S, Labrecque O, Roy JP, Fecteau G. Risk factors associated with transmission of *Mycobacterium avium subsp. paratuberculosis* to calves within dairy herd: a systematic review. *J Vet Intern Med*. (2012) 26:32–45. doi: 10.1111/j.1939-1676.2011.00854.x
37. Timsit E, Dendukuri N, Schiller I, Buczinski S. Diagnostic accuracy of clinical illness for bovine respiratory disease (BRD) diagnosis in beef cattle placed in feedlots: a systematic literature review and hierarchical Bayesian latent-class meta-analysis. *Prev Vet Med*. (2016) 135:67–73. doi: 10.1016/j.prevetmed.2016.11.006
38. de Boer MW, LeBlanc SJ, Dubuc J, Meier S, Heuwieser W, Arlt S, et al. Invited review: systematic review of diagnostic tests for reproductive-tract infection and inflammation in dairy cows. *J Dairy Sci*. (2014) 97:3983–99. doi: 10.3168/jds.2013-7450
39. Glanville J, Wood H, Arber M, Varley D, Frampton G, Brazier H. *Technical Manual for Performing Electronic Literature Searches in Food and Feed Safety*. (2013). Available online at: https://www.yhec.co.uk/yhec-content/uploads/2015/03/Glanville-EFSA-search-manual-Sept-2013.pdf (accessed October 3, 2019).
40. Wood H, O'Connor A, Sargeant J, Glanville J. Information retrieval for systematic reviews in food and feed topics: a narrative review. *Res Synth Methods*. (2018) 9:527–39. doi: 10.1002/jrsm.1289
41. Grindlay DJ, Brennan ML, Dean RS. Searching the veterinary literature: a comparison of the coverage of veterinary journals by nine bibliographic databases. *J Vet Med Educ*. (2012) 39:404–12. doi: 10.3138/jvme.1111.109R
42. Paez A. Gray literature: an important resource in systematic reviews. *J Evid Based Med*. (2017) 10:233–40. doi: 10.1111/jebm.12266
43. Sargeant JM, Thompson A, Valcour J, Elgie R, Saint-Onge J, Marcynuk P, et al. Quality of reporting of small animal clinical trials and associations with treatment effects. *J Vet Intern Med*. (2010) 24:44–50. doi: 10.1111/j.1939-1676.2009.0386.x
44. Sargeant JM, Elgie R, Valcour J, Saint-Onge J, Thompson A, Marcynuk P, et al. Methodological quality and completeness of reporting in clinical trials conducted in livestock species. *Prev Vet Med*. (2009) 91:107–15. doi: 10.1016/j.prevetmed.2009.06.002
45. Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, et al. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. *J Clin Epidemiol*. (2013) 66:158–72. doi: 10.1016/j.jclinepi.2012.01.012
46. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. (2019) 366:l4898. doi: 10.1136/bmj.l4898
47. Moura CAA, Totton SC, Sargeant JM, O'Sullivan TL, Linhares DCL, O'Connor AM. Evidence of improved reporting of swine vaccination trials in the post-REFLECT statement publication period. *J Swine Health Prod*. (2019) 27:265–77.
48. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. *BMJ*. (2016) 355:i4919. doi: 10.1136/bmj.i4919
49. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. *The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomized Studies in Meta-Analyses*. Available online at: http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm (accessed October 3, 2019).
50. Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epidemiol*. (2012) 65:163–78. doi: 10.1016/j.jclinepi.2011.05.008
51. Broen MP, Braaksma MM, Patijn J, Weber WE. Prevalence of pain in Parkinson's disease: a systematic review using the Modified QUADAS Tool. *Move Disord*. (2012) 27:480–4. doi: 10.1002/mds.24054
52. Bakbergenuly I, Hoaglin DC, Kulinskaya E. Pitfalls of using the risk ratio in meta-analysis. *Res Synth Methods*. (2019) 10:398–419. doi: 10.1002/jrsm.1347
53. Rothman K, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins (2008).
54. Cooper H, Hedges LV, Valentine JC. *The Handbook of Research Synthesis and Meta-Analysis*. 3rd ed. New York, NY: Russell Sage Foundation (2019).
55. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. West Sussex: John Wiley and Sons Ltd. (2009). doi: 10.1002/9780470743386
56. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. (1986) 7:177–88. doi: 10.1016/0197-2456(86)90046-2
57. Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods*. (2016) 7:55–79. doi: 10.1002/jrsm.1164
58. Deeks JJ, Higgins JPT, Altman DG. Chapter 10: analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.0*. (2019). Available online at: www.training.cochrane.org/handbook (accessed November 30, 2019).
59. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat Med*. (2001) 20:3875–89. doi: 10.1002/sim.1009
60. Sidik K and Jonkman JN. A simple confidence interval for meta-analysis. *Stat Med*. (2002) 21:3153–9. doi: 10.1002/sim.1262
61. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc*. (2009) 172:137–59. doi: 10.1111/j.1467-985X.2008.00552.x
62. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. (2003) 327:557–60. doi: 10.1136/bmj.327.7414.557
63. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. (1997) 315:629–34. doi: 10.1136/bmj.315.7109.629
64. Brace S, Taylor D, O'Connor AM. The quality of reporting and publication status of vaccines trials presented at veterinary conferences from 1988 to 2003. *Vaccine*. (2010) 28:5306–14. doi: 10.1016/j.vaccine.2010.05.041
65. Snedeker KG, Totton SC, Sargeant JM. Analysis of trends in the full publication of papers from conference abstracts involving pre-harvest or

abattoir-level interventions against foodborne pathogens. *Prev Vet Med*. (2010) 95:1–9. doi: 10.1016/j.prevetmed.2010.02.015
66. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE*. (2008) 3:e3081. doi: 10.1371/journal.pone.0003081
67. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. (2011) 343:d4002. doi: 10.1136/bmj.d4002
68. Wisener LV, Sargeant JM, O'Connor AM, Faires MC, Glass-Kaastra SK. The evidentiary value of challenge trials for three pre-harvest food safety topics: a systematic assessment. *Zoon Public Health*. (2014) 61:449–76. doi: 10.1111/zph.12083
69. Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *J Clin Epidemiol*. (2005) 58:894–901. doi: 10.1016/j.jclinepi.2005.01.006
70. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics*. (1994) 50:1088–1101. doi: 10.2307/2533446
71. Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. *J Clin Epidemiol*. (2000) 53: 207–16 doi: 10.1016/S0895-4356(99)00161-4
72. Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *BMJ. (*2000) 320:1574–7. doi: 10.1136/bmj.320.7249.1574
73. Duval S, Weinhandl E. *Correcting for Publication Bias in the Presence of Covariates*. Agency for Healthcare Research and Quality (US) (2011). Available online at: https://www.ncbi.nlm.nih.gov/books/NBK84228 (accessed November 30, 2019).
74. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J. Clin. Epidemiol*. (2009) 62:1006–12. doi: 10.1016/j.jclinepi.2009.06.005
75. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med*. (2009) 6:e1000100. doi: 10.1371/journal.pmed.1000100
76. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. (2008) 336:924–6. doi: 10.1136/bmj.39489.470347.AD
77. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. (2011) 64:401–6. doi: 10.1016/j.jclinepi.2010.07.015
78. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol*. (2011) 64:407–15. doi: 10.1016/j.jclinepi.2010.07.017
79. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence–publication bias. *J Clin Epidemiol*. (2011) 64:1277–82. doi: 10.1016/j.jclinepi.2011.01.011
80. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence–imprecision. *J Clin Epidemiol*. (2011) 64:1283–93. doi: 10.1016/j.jclinepi.2011.01.012
81. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE Working Group. GRADE guidelines: 7. Rating the quality of evidence–inconsistency. *J Clin Epidemiol*. (2011) 64:1294–302. doi: 10.1016/j.jclinepi.2011.03.017
82. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE Working Group. GRADE guidelines: 8. Rating the quality of evidence–indirectness. *J Clin Epidemiol*. (2011) 64:1303–10. doi: 10.1016/j.jclinepi.2011.04.014
83. Sullivan SL, Whittem T, Morley PS, Hinchcliff KW. A systematic review and meta-analysis of the efficacy of furosemide for exercise-induced pulmonary haemorrhage in Thoroughbred and Standardbred racehorses. *Equine Vet J*. (2015) 47:341–9. doi: 10.1111/evj.12373
84. Baltzell P, Newton H, O'Connor AM. A critical review and meta-analysis of the efficacy of whole-cell killed *Tritrichomonas foetus* vaccines in beef cattle. *J Vet Intern Med*. (2013) 27:760–70. doi: 10.1111/jvim.12112
85. Wilhelm B, Rajić A, Parker S, Waddell L, Sanchez J, Fazil A, et al. Assessment of the efficacy and quality of evidence for five on-farm interventions for *Salmonella* reduction in grow-finish swine: a systematic review and meta-analysis. *Prev Vet Med*. (2012) 107:1–20. doi: 10.1016/j.prevetmed.2012.07.011
86. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open*. (2017) 7:e012545. doi: 10.1136/bmjopen-2016-012545
87. Lau J. Editorial: systematic review automation thematic series. *Syst Rev*. (2019) 8:70. doi: 10.1186/s13643-019-0974-z
88. O'Connor AM, Tsafnat G, Gilbert SB, Thayer KA, Shemilt I, Thomas J, et al. Still moving toward automation of the systematic review process: a summary of discussions at the third meeting of the International Collaboration for Automation of Systematic Reviews (ICASR). *Syst Rev*. (2019) 8:57. doi: 10.1186/s13643-019-0975-y
89. O'Connor AM, Tsafnat G, Thomas J, Glasziou P, Gilbert SB, Hutton B. A question of trust: can we build an evidence base to gain trust in systematic review automation technologies? *Syst Rev*. (2019) 8:143. doi: 10.1186/s13643-019-1062-0
90. Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *Syst Rev*. (2019) 8:163. doi: 10.1186/s13643-019-1074-9
91. Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, et al. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *JAMA*. (1998) 280:278–80. doi: 10.1001/jama.280.3.278
92. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction-the why, what, when, and how. *J Clin Epidemiol*. (2017) 91:23–30.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 14 February 2020
doi: 10.3389/fvets.2020.00065

Economic Impact of a Bluetongue Serotype 8 Epidemic in Germany

Jörn Gethmann, Carolina Probst and Franz J. Conraths*

Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Institute of Epidemiology, Greifswald-Insel Riems, Germany

Background and Objectives: Germany was affected by Bluetongue virus serotype 8 (BTV-8) from 2006 to 2009 and recorded new cases since December 2018. We assessed the economic impact of the epidemic from the first cases in 2006 until 2018. Direct costs include production losses, animal deaths, and veterinary treatment. Indirect costs include surveillance, additional measures for animal export, disease control (preventive vaccination and treatment with insecticides), vector monitoring, and administration.

Methodology: To estimate the financial impact of BTV-8 on different species and production types at the animal level, we performed a gross margin analysis (GMA) for dairy and beef cattle, and sheep. To estimate the impact on the national level, we used a modified framework described by Rushton et al. (1) and applied a methodology described by Bennett (2). Both the GMA and the economic model on national level were implemented in Excel and the Excel Add-in @Risk. The tools, which are widely applicable, also for other diseases, are made available here.

Results: The financial impact of a BTV-8 infection at the animal level was estimated at 119–136 Euros in dairy cattle, at 27 Euros in beef cattle, and at 74 Euros in sheep. At the national level, the impact of the BTV-8 epidemic ranged between 157 and 203 million Euros (mean 180 million Euros). This figure consisted of 132 (73%) and 48 (27%) million Euros for indirect and direct costs. Indirect costs included 89 million Euros (67%) for vaccination, 18 million Euros (14%) for insecticide treatment, 15 million Euros (11%) for diagnostic testing of animals dispatched for trade, 8 million Euros (6%) for monitoring and surveillance, and 3 million Euros (2%) for administration. The highest costs were induced by a compulsory vaccination campaign in 2008 (51 million Euros; 28% of the total costs) and the disease impact on cattle in 2007 (30 million Euros; 17%).

Discussion: We compare the outcome of our study with economic analyses of Bluetongue disease in other countries, and discuss the suitability of GMA and the developed tools for a wider application in veterinary economics.

Keywords: Bluetongue disease, cattle, sheep, economy, gross margin analysis, disease control, costs

OPEN ACCESS

Edited by:
Andres M. Perez, University of Minnesota Twin Cities, United States

Reviewed by:
Armin R. W. Elbers, Wageningen Bioveterinary Research (WBVR), Netherlands
Christie Mayo, Colorado State University, United States
Claudia Schulz, University of Veterinary Medicine Hannover, Germany

****Correspondence:***
Jörn Gethmann
joern.gethmann@fli.de

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *26 November 2019*
Accepted: *27 January 2020*
Published: *14 February 2020*

Citation:
Gethmann J, Probst C and Conraths FJ (2020) Economic Impact of a Bluetongue Serotype 8 Epidemic in Germany. Front. Vet. Sci. 7:65. doi: 10.3389/fvets.2020.00065

INTRODUCTION

Bluetongue (BT) is a non-contagious infectious disease of domestic and wild ruminants caused by the BT virus (BTV), which belongs to the genus *Orbivirus* within the family *Reoviridae*. The virus is transmitted by biting midges of the genus *Culicoides*. At least 27 BTV serotypes have so far been detected worldwide (3). All ruminants are susceptible, although clinically apparent disease is most often reported in sheep. During the BTV-8 epidemic in 2006–2010 in Germany, clinical symptoms were observed in both, cattle and sheep. The most frequently reported signs were fever, weight loss, apathy, erosions of the oral mucosa, salivation, dysphagia, oedema of the head and lips, lameness, reduced milk yield and abortions (4–7). Effects on animal production in sheep and cattle for different BT serotypes and settings have been reviewed by Rushton and Lyons (8).

In August 2006, BTV-8 emerged for the first time almost simultaneously in Belgium, France, Germany, and the Netherlands (6, 9, 10). The disease hit an immunologically naïve and thus highly susceptible population. In Germany, the disease was first detected in late August 2006 (11). By the end of 2006, a total of 890 BTV-8 cases had been recorded in four German federal states and reported to the German Animal Disease Notification System (https://tsn.fli.de; public site TSIS: https://tsis.fli.de).

To determine the distribution and spread of BTV-8, the European Commission issued instructions for monitoring and surveillance in the member states of the European Union (SANCO/10581/2006 Rev 4). These included serological surveys and testing sentinel animals for antibodies to detect potential new cases as early as possible in 2007. Based on this working document, the European Commission regulation (EC) No 1266/2007 was launched, which established harmonized disease control measures, including preventive vaccination, a sentinel program, vector monitoring and monitoring in wild ruminants.

The following measures were initiated in Germany (**Figure 1**): (i) a cross-sectional study (February to April 2007) within the 150 km restriction zone to assess the prevalence of BTV-8 infections in cattle and sheep (12); (ii) a sentinel program to detect the re-occurrence of BT, during which ∼150 animals from 10 to 15 farms were monthly tested for antibodies to BTV-8 in each federal state; (iii) wildlife monitoring (2007 until today); (iv) vector monitoring to obtain information on the distribution and seasonal activity of potential BTV-8 vectors. Further disease control measures included the improvement of biosafety at the farm level, treatment of animals and stables with insecticides, and the testing of animals dispatched for trade.

Despite these measures BTV-8 re-occurred in May 2007. The disease spread over wide parts of Germany and affected more than 20,000 farms, causing the death of animals and substantial production losses, especially in sheep (11, 13–16).

As soon as commercial vaccines against BTV-8 had become available, they were tested for safety and efficacy (17, 18). Germany then initiated a country-wide mandatory vaccination campaign for cattle, sheep and goats, which started in May 2008. During 2008 and 2009, the number of outbreaks decreased sharply. Especially in the affected regions, farmers perceived vaccination positively. They realized that it prevented output losses and allowed trade and animal movements without restrictions. However, when vaccination became voluntary in 2010, farmers' willingness to vaccinate against BTV-8 was estimated at only 43% for cattle and 34% for sheep (19). Nevertheless, the BTV-8 epidemic subsided, so that Germany was declared officially free from BTV on 15 February 2012.

In August 2015, BTV-8 re-emerged in France (20) and in 2017 also in Switzerland (21). Compared to the strain that had circulated from 2006 to 2009, the current BTV-8 strain seemed to be less virulent, although the genome of the virus had remained stable (22, 23). In November 2017, a second BTV-serotype (BTV-4) was introduced to France (24). In view of the situation in the neighboring countries, the German federal state of Baden-Wuerttemberg encouraged farmers to vaccinate cattle, sheep and goats against both, BTV-8 and BTV-4, also by providing financial support. However, vaccination is not mandatory, and since the costs have to be borne mainly by farmers, the vaccine coverage was only about 25% by the end of 2018 (25). On 12 December 2018, two BTV-8-positive cattle were detected in Baden-Wuerttemberg as part of routine BTV surveillance (animals were tested by PCR and serological tests on December 6th; the test results were confirmed by the national reference laboratory; the outbreaks were recorded in the German animal disease notification system on 12th December 2018). The animals were clinically healthy. Again, monitoring was intensified and consigned animals could only be moved from this region to areas not under restriction, if the animals had been vaccinated against BTV-8 or tested for BT with a negative result (according to Article 8 of Commission Regulation (EC) No 1266/2007 serological or agent identification test).

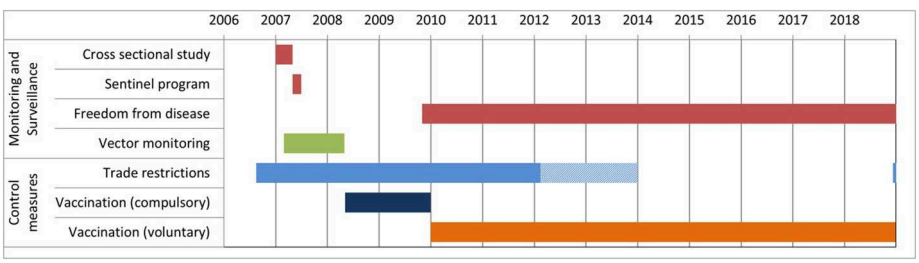


FIGURE 1 | Timeline of official monitoring, surveillance, and control measures regarding BTV-8 in Germany.

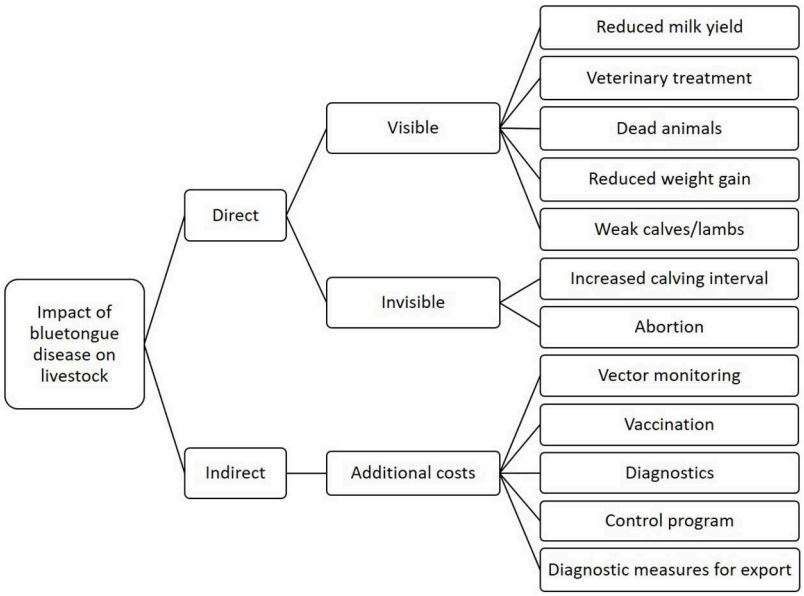


FIGURE 2 | Impact of Bluetongue disease on cattle and sheep.

There are ∼11.8 million cattle in Germany (thereof about 40% dairy) and 1.6 million sheep (as of Mai 2019, Federal statistical office; https://www.destatis.de/). Livestock farming is the main source of income in agriculture in the country (26). Especially in the light of the re-emergence of BTV-8 and a potential future introduction of other BTV serotypes, namely BTV-4, disease contingency plans are evaluated, also from an economic point of view.

The purpose of this study was to carry out an ex-post economic impact analysis for BTV-8 in Germany for the years 2006–2018. The aim is to provide stakeholders and decision makers with a transparent evaluation of the potential benefits of preventing and controlling a vector-borne disease in livestock. The tools developed for the assessments and calculations are made available with this publication, so that they can be applied to other diseases and scenarios.

MATERIALS AND METHODS

To calculate direct losses on the animal level, we used a gross margin analysis (GMA). To estimate the economic impact of BT on the national level, we applied a modified framework previously described by Rushton et al. (1, 27) (**Figure 2**) and a standardized method described by Bennett (2). The method can be adapted to BT as described elsewhere (28). Both, the GMA and the economic model run at the national level were implemented in a stochastic-deterministic spreadsheet in Excel version 2019 (Microsoft® GmbH, Unterschleißheim, Germany) and @Risk 7.0.0 (Palisade Corporation, Newfield, NY, USA). @RISK is an add-in to Microsoft Excel that allows analyzing risks using Monte Carlo simulation (https://www.palisade.com/). The spreadsheets for both, the GMA and the economic model, as well as user manuals, are provided in **Supplementary Material**.

The national average of herd performance as well as epidemiological and economic data were collected from Eurostat (https://ec.europa.eu/eurostat), the official German animal disease notification system (*TierSeuchenNachrichten-System*, https://tsn.fli.de/; public site: https://tsis.fli.de), the Federal statistical office (*Statistisches Bundesamt*, https://www.destatis.de/), the Identification and Information System for Animals (*Herkunftssicherungs- und Informationssystem für Tiere*, https://www.hi-tier.de/), and the Federal Office for Agriculture and Food (*Bundesanstalt für Landwirtschaft und Ernährung*, https://www.ble.de/). Other input parameters were obtained from the Bavarian State Agency for Agriculture (*Bayerische Landesanstalt für Landwirtschaft*, https://www.stmelf.bayern.de/idb/default.html), the German Association for Technology and Structures in Agriculture (*Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V.*, https://www.ktbl.de/), the animal health services of the federal states (*Tiergesundheitsdienste*), and the animal disease compensation funds of the federal states (*Tierseuchenkassen*). Values for the surveillance costs were obtained from reports of the federal states to the German Federal Ministry for Food and Agriculture and the annual applications of the Federal Ministry to the European Commission for co-financing animal disease control and surveillance (https://ec.europa.eu/food/funding/animal-health/national-veterinary-programmes_en). Parameters with fluctuation (e.g., milk yield, milk price) were resampled from known values with 10,000 iterations. Details are given in **Supplementary Tables S1–S3**. All monetary values are expressed in Euros.

Economic Impact at the Animal Level

Direct costs at the animal level, i.e., production losses due to clinical illness (dC_{py}), were estimated by calculating the difference between the gross margin (GM) of a healthy animal

and the GM of a clinically ill animal. The GM was calculated per year (2006–2018) separately for dairy cattle (GM_{Dy}), beef cattle (GM_{Fy}), and sheep (GM_{Sy}). Since infected calves and heifers were rarely reported to show clinical signs, we focused on adult animals. The GMA was performed according to standard procedures (https://www.stmelf.bayern.de/idb/default.html). To calculate the impact of BTV-8 at the animal level, we included the value of lost animal production with the disease-related intervention costs. Details of the GMA are provided in **Supplementary Tables S1A** (GM_{Dy}), **S1B** (GM_{Fy}), **S1C** (GM_{Sy}). For example, GM_{Dy} was calculated as follows:

$$GM_{Dy} = (R_{miy} + R_{any} + R_{may}) - (VC_{ry} + VC_{fey} + VC_{cry} + VC_{vy} + VC_{way} + VC_{insy} + VC_{macy} + VC_{laby} + VC_{misy})$$

i.e., including the revenues (R) for selling milk (R_{miy}), animals (R_{any}), and manure (R_{may}), and the variable costs (VC) for restocking (VC_{ry}), feed (VC_{fey}), calf rearing (VC_{cry}), veterinary treatment (VC_{vy}), water/electricity (VC_{way}), insemination (VC_{insy}), machines (VC_{macy}), hired labor (VC_{laby}) and miscelaneous (VC_{misy}).

For dairy cattle, parameters influenced by BTV-8 are associated with reduced fertility (prolonged calving interval due to abortion and stillbirth), lower milk yield, and costs for restocking (assuming the goal of maintaining the herd size constant despite increased mortality of calves and cows as the reference). In beef cattle, the main impact of BTV-8 was due to a drop in feed intake during the first days of disease, resulting in reduced daily weight gain and thus in a prolonged fattening period (assuming the goal of reaching the usual slaughter weight as the reference). For sheep, losses were mainly attributed to reduced revenues for selling animals (due to reduced slaughter weight) and the need to purchase new ewes (increased replacement rate resulting from death of ewes and lamb losses). In all animal species, the disease increased the costs for veterinary treatment. Details on the influence of BTV-8 on the GM are provided in **Supplementary Tables S2A–C**.

Economic Impact on National Level

The results of the GMA were extrapolated from the animal level to the national population level separately for each species (cattle, sheep), production type (dairy, meat), and year (2006–2018). Details on the economic model are provided in **Supplementary Table S3**. The economic impact (net total costs) at the national level (C_{BT}) includes direct (DC_y) and indirect costs (IC_y) and was calculated as follows:

$$C_{BT} = \sum_{y=2006}^{2018} DC_y + IC_y$$

Direct Costs

Direct costs at the national level per year (DC_y) include production losses due to clinical illness (DCc_y) and the value of the animals that succumbed to the disease (DCd_y):

$$DCy = DCc_y + DCd_y$$

Production losses

To estimate the production losses caused by clinical illness ($DCDc_y$, $DCFc_y$, $DCSc_y$), we multiplied the number of animals that had developed clinical signs with the average direct costs per animal (dCD_y, dCF_y, dCS_y), which had previously been calculated in the GMA (which also included the veterinary costs). The number of animals that developed clinical signs were estimated by multiplying the number of newly infected cattle (nCi_y) and sheep (nSi_y) with the estimated morbidity for cattle (rcc) and sheep (rsc), and in case of cattle, with the proportion of dairy (pd) or beef (pf) cattle in the total cattle population.

$$DCDc_y = nCi_y * rcc * pd/100 * dC_{py}$$
$$DCFc_y = nCi_y * rcc * pf/100 * dC_{py}$$
$$DCSc_y = nSi_y * rsc * dC_{py}$$

For cattle (rcc), morbidity was estimated to range between 5 and 15%, and for sheep (rsc), between 15 and 25%. The number of newly infected cattle and sheep per year were estimated as follows:

$$nCi_y = \frac{I_{sy}}{100} * ncz_y$$
$$nSi_y = \frac{I_{sy}}{100} * nsz_y$$

where I_{sy} is the yearly incidence per species (cattle, sheep), and ncz_y and nsz_y the number of cattle and sheep in the restriction zones.

To estimate I_{sy} for the year 2006, we used the results of a cross-sectional study performed in early 2007 (12). Since this study had revealed an underreporting of outbreaks (affected farms), we assumed that the relation between the numbers of officially reported outbreaks (P_y) and I_{sy} remained constant over the years. I_{sy} was therefore estimated as follows:

$$I_{sy} = \frac{I_{2006}}{P_{2006}} * P_y$$

To estimate the numbers of newly infected cattle (nCi_y) and sheep (nSi_y), we multiplied the number of cattle (ncz_y) or sheep (nsz_y) in the restriction zones with the respective incidence. Regarding the number of animals in the restriction zones, we used the numbers of cattle and sheep kept in the BT-affected federal states for the year 2006, and the whole German cattle population for the years 2007–2011. Since 2012, no restriction zones for BT had remained.

Animal losses

To estimate the value of animals that succumbed to disease (DCd), the number of dead animals was multiplied with the compensation paid by the animal disease compensation fund for cattle (vc_y) and sheep (vs_y). The numbers of dead cattle (nCd_y) or sheep (nSd_y), respectively, were estimated based on data provided by the animal disease compensation funds and animal health services for the year 2007, assuming that the relation between newly infected (nCi_y, nSi_y) and dead animals (nCd_y, nSd_y) in

2007 (mean case-fatality ratio) remained constant throughout the years. Compensation includes the common value of the animals and for cattle also the disposal costs.

Indirect Costs

Indirect costs at the national level per year (IC_y) include the costs for all legal provisions successively implemented to control BTV-8, including surveillance (ICS_y), measures for animal export (ICE_y), treatment with insecticides (ICI_y), vaccination (ICV_y), vector monitoring (ICM_y) and administration time for establishing restriction zones and reporting (ICA_y):

$$IC_y = ICS_y + ICE_y + ICI_y + ICV_y + ICM_y + ICA_y$$

Surveillance

The costs for BT surveillance according to SANCO/10581/2006 Rev 4 and Commission Regulation (EC) No 1266/2007 (ICS_y) include the costs for the cross-sectional study performed in winter 2007, the sentinel program (early detection) performed in early 2007, and the BT monitoring for disease detection performed between 2007 and 2018. For cattle, sheep and goats, respectively, the costs for BT monitoring were calculated as follows:

$$\begin{aligned} ICS_y &= nfs_y * (ct_f + cp_f) + ns_{sy} * csa_s + n_{ELI} * c_{ELI} \\ &\quad + n_{PCR} * c_{PCR} \end{aligned}$$

where nfs_y is the number of tested farms, ct_f and cp_f the travel and personnel costs per tested farm, ns_{sy} and csa_s the number of samples and sampling costs per species, n_{ELI} and n_{PCR} the number of samples tested by enzyme-linked immunosorbent assay (ELISA) or polymerase chain reaction (PCR), and c_{ELI} and c_{PCR} the respective costs for testing.

The numbers of tested farms and animals, as well as the costs for sampling and laboratory analysis were retrieved from the national applications for co-financing (Commission Decisions 2007/20/EC, 2008/655/EC, 2009/883/EC, 2010/712/EU, 2011/807/EU, 2012/282/EU, 2012/761/EU, 2013/722/EU, 2014/925/EU, 2014/288/EU, 2015/2444/EU, and 2016/969/EU). These documents are published at https://ec.europa.eu/food/funding/animal-health/national-veterinary-programmes_en.

Travel costs for official veterinarians (ct_f) were estimated as follows:

$$ct_f = k * d * 2$$

where k is the fee per km and d the average distance between the veterinary office and a farm.

Personnel costs for official veterinarians (cp_f) were estimated as follows:

$$cp_f = ts_f * cp_h$$

where ts_f is the time spent on the farm and cp_h the average personnel costs of an official veterinarian per hour.

Measures for animal export

Following Commission Regulation (EC) No 1266/2007, Germany was between 2006 and 2008 requested to confirm that all animals intended for movement to BT-free EU member states or export to third countries were negative either in a BT-specific ELISA or by PCR. These additional costs for movement or export testing (ICE_y) had to be borne by the farmers and were estimated as follows:

$$ICE_y = (nce_y * pe_y * cet_y) + (nse_y * pe_y * cet_y)$$

where nce_y and nse_y are the numbers of cattle and sheep exported per year, pe_y the proportion of animals that were exported to BT-free countries and had therefore to be tested, and cet_y the test costs per animal. The costs caused by animal movements within the country were not taken into account, since almost all regions of Germany were part of a single restriction zone from 2007 until 2012.

Insecticide treatment

The costs for treatment with insecticides (ICI_y) were calculated as follows:

$$\begin{aligned} ICI_y &= (ncz_y * ci_y + ncfz_y * ci_{fy}) * pic_y + (nisf_y * s_f \\ &\quad * ci_y) + (nisf_y * ci_{fy}) \end{aligned}$$

where ncz_y is the number of cattle in restriction zones, ci_y the costs for insecticides, $ncfz_y$ the number of cattle farms in restriction zones, ci_{fy} the personnel costs per treated farm, pic_y the proportion of cattle (animals and farms) treated with insecticides, $nisf_y$ the number of infected sheep farms and s_f the mean number of sheep per farm.

Vaccination

Depending on the animal species and the vaccine, two or three injections were required for a complete basic immunization. After that, the animals had to be re-vaccinated once a year. To calculate vaccination costs at the animal level for cattle ($ICVc_y$) and sheep ($ICVs_y$), we accounted for the costs per vaccinated animal and the costs per vaccinated farm. The costs per vaccinated animal were estimated by multiplying the number of vaccinations (cattle $nvdc_y$ or sheep nvs_y) with the sum of the costs for the vaccine (cattle cvc_d or sheep cvs_d) and vaccination per immunization dose (cattle $cvac_d$ or sheep $cvas_d$). Since vaccines had to be applied by a veterinarian, vaccination costs at the farm level include the numbers of vaccinated farms (cattle $nvcf_y$ or sheep $nvsf_y$) and the herd fee charged by the veterinarian (cattle cc_f or sheep cs_f) according to the following equations:

$$\begin{aligned} ICVc_y &= nvdc_y * (cvc_d + cvac_d) + (nvcf_y * cc_f) \\ ICVs_y &= nvs_y * (cvs_d + cvas_d) + (nvsf_y * cs_f) \end{aligned}$$

Vector monitoring

The costs for vector monitoring (ICM_y) account for the costs for vector traps and data loggers ($cvmt_y$), trap management ($cvmm_y$), and entomological tests (cvme):

$$ICM_y = nvt_y * (cvmt_y + cvmm_y) + nvme_y * cvme_y$$

where nvt_y are the number of vector traps, and $nvme_y$ the number of entomological tests.

Administration

The administration costs include the costs for epidemiological investigations (farm visits), for the establishment of restriction zones and the time required for reporting. Personnel and travel costs for farm visits of official veterinarians were also taken into account, so that calculations were performed according to the following equation:

$$ICA_y = tP_y * d * k * 2 + (t_f * p_h)$$

where tP_y ist the prevalence of BT (total number of affected cattle and sheep farms per year), d the average distance between the veterinary office and the affected farm, k the fee charged per driven km, t_f the average time spent per farm, and p_h the average personnel costs for an official veterinarian per working hour.

Sensitivity Analysis

We analyzed the effect of the input variables where we entered a distribution on the output mean (sensitivity analysis). The sensitivity analysis was performed in @Risk 7.0.0 (Palisade) with a one-at-a-time method (29, 30), where each variable is analyzed separately. The sensitivity analysis was carried out as follows (https://kb.palisade.com/index.php?pg=kb.page&id=248): (1) all iterations are ranked by ascending values of the input; (2) the ranked iterations are attributed to 10 bins (in this case with 10,000 iterations, each bin contains 1000 values); (3) the mean of the output values of each bin is computed; (4) 10 output means from the bins are compared. The lowest output mean gets the number at the left edge; the highest of the 10 output means receives the number at the right edge.

Finally, the results of the sensitivity analysis were assessed qualitatively using a tornado plot, which shows how the mean of the model varies over the range of each input variable.

RESULTS

Economic Impact on Animal Level

For the dairy sector, direct costs ranged between 119 and 136 Euros per infected animal, depending on the milk price (**Table 1**). Most of the costs resulted from the need to restock (99 Euros/animal), veterinary treatment (26 Euros) and production losses (24 and 18 Euros less output for milk and calf sales, respectively). In the beef sector, direct costs amounted to 27 Euros per animal on average. They were mainly attributable to the prolonged fattening period. For sheep, direct costs were estimated at 74 Euros per animal on average. They were mainly due to reduced revenues for lamb sales (59 Euro per infected ewe) and veterinary treatment, especially after abortions (10 Euros/animal) (data not shown).

Economic Impact on National Level

The net total costs of the BTV-8 epidemic in Germany, including prevention and control measures over the last 13 years (2006–2018), ranged between 157 and 203 million Euros (mean 180.4 million Euros, standard deviation 6 million) (**Table 2**). This figure includes on average 132.1 (73%) million Euros indirect and 48.3 (27%) million Euros direct costs.

Mean indirect costs included 106.5 million Euros for disease control measures (vaccination and insecticide treatment, 59% of the net total costs), 14.9 million Euros for additional measures relating to export (12.3 million only for cattle), 7.9 million Euros for BT monitoring and surveillance (including 1.2 million Euros for vector monitoring in 2007 and 2008), and 2.8 million Euros for administration. Disease control measures consisted of 88.6 million Euros for vaccination (74 and 14 million Euros for cattle and sheep, respectively) and 18.0 million Euros for treatment with insecticides (16.9 and 1.1 million Euros for cattle and sheep) (see **Tables 2**, **3**).

Mean direct costs mainly arose in the cattle sector (37.4 million Euros, 21% of the net total costs) (**Table 2**). In the sheep sector, they amounted to 10.9 million Euros (6%). Direct costs were highest in 2007, when they reached 39.8 million Euros (29.7 million in cattle, 10.1 million in sheep) (**Table 3**). In 2007, the animal compensation funds paid for 10,240 cattle and 33,233 sheep that prematurely died due to BTV-8 infection. This corresponds to a mortality ratio of 0.081 for cattle and 1.4 for sheep. The compensation paid per animal was 1,500–1,900 Euros for cattle and 120–170 Euros for sheep, including rendering costs.

TABLE 1 | Direct costs of a BTV-8 infection per animal, with minimum, mean, maximum, 5 and 95% percentiles in million Euros.

Gross margin	Minimum	Mean	Maximum	5%	95%
Dairy 2006	78	122	515	92	173
Dairy 2007	78	129	928	94	194
Dairy 2008	90	136	614	102	198
Dairy 2009	79	119	391	91	164
Beef (2006–2009)	14	27	40	22	33
Sheep (2006–2009)	42	74	104	60	88

TABLE 2 | Minimum, maximum and mean net costs (in million Euros) of BTV-8 in Germany from 2006 to 2018) with 5% and 95% percentiles.

Cost factor	Minimum	Mean	Maximum	5%	95%
Net total costs	157.002	180.406	202.995	169.915	191.056
Total direct costs	37.091	48.313	60.842	42.403	54.554
Direct costs cattle	27.756	37.449	50.226	31.797	43.372
Direct costs sheep	7.482	10.864	14.893	8.802	12.965
Total indirect costs	115.836	132.092	149.548	123.500	140.840
Vaccination cattle	64.148	74.497	85.153	67.521	81.543
Vaccination sheep	12.996	14.064	15.187	13.450	14.682
Insecticide treatment cattle	12.518	16.894	21.176	14.476	19.269
Insecticide treatment sheep	805	1.078	1.349	926	1.233
Export measures cattle	7.153	12.263	20.769	8.773	17.047
Export measures sheep	1.782	2.627	3.431	2.183	3.069
Monitoring and surveillance	7.562	7.882	8.225	7.700	8.070
Administration	1.755	2.788	3.917	2.171	3.421

Totals and subtotals are indicated by gray shading.

TABLE 3 | Mean net total costs of BTV-8 in Germany per year from 2006 to 2018 (in million Euros).

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Net total costs	9.250	59.105	66.810	27.022	10.441	5.910	1.453	0.177	0.076	0.162	2.358	2.192	1.528
Total direct costs	1.863	39.765	6.664	0.022	0	0	0	0	0	0	0	0	0
Direct costs cattle	1.461	29.661	6.308	0.020	0	0	0	0	0	0	0	0	0
Direct costs sheep	0.402	10.105	0.356	0.002	0	0	0	0	0	0	0	0	0
Total indirect costs	7.387	19.339	60.146	27.001	10.441	5.910	1.453	0.177	0.076	0.162	2.358	2.192	1.528
Insecticide treatment cattle	2.893	7.959	2.063	1.021	1.002	0.982	0.974	0	0	0	0	0	0
Insecticide treatment sheep	0.040	1.002	0.035	0	0	0	0	0	0	0	0	0	0
Vaccination cattle	0	0	44.530	17.284	7.916	4.429	0.305	0.031	0.001	0.001	1.462	1.691	1.199
Vaccination sheep	0	0	6.777	5.549	1.358	0.344	0.032	0.003	0	0	0.740	0.386	0.209
Export measures	3.500	5.047	3.929	2.415	0	0	0	0	0	0	0	0	0
Administration	0.190	2.273	0.324	0.001	0	0	0	0	0	0	0	0	0
Monitoring and surveillance	0.765	3.058	2.488	0.730	0.165	0.155	0.142	0.143	0.075	0.161	0.156	0.115	0.119

Totals and subtotals are indicated by gray shading.

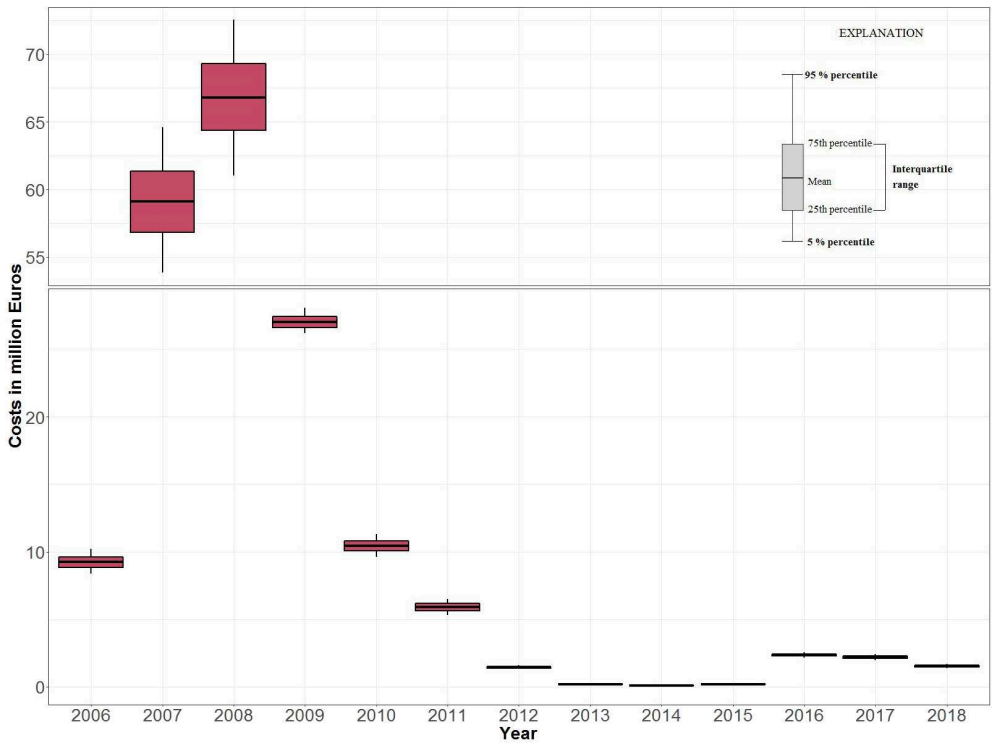


FIGURE 3 | Boxplots of the net total costs of BTV-8 in Germany (in million Euros) with mean values, 25%–75% (box) and 5%–95% percentiles.

This corresponds to total compensation payments of 17.3 and 4.2 million Euros for cattle and sheep in Germany.

The yearly costs were highest in 2008 and 2007, with 66.8 (37% of net total costs) and 59.1 million (32%) Euros, respectively (**Table 3** and **Figure 3**). After peaking in 2008, they gradually dropped from 27.0 million (2009) to 74 thousand Euros (2014). In 2015, they started to increase again and reached 1.5 million Euros in 2018 (**Table 3** and **Figure 4**).

In 2007, the total costs were mainly attributable to direct costs caused by BTV-8 infections. In 2008, total costs accrued mainly from vaccination (51.3 million Euros, thereof 44.5 million Euros for cattle, 25% of the total costs) (**Table 3**). In 2009 and 2010, vaccination of cattle cost 17.3 and 7.9 million Euros (10% and 5%), respectively. Since 2010, the costs for voluntary vaccination had to be borne by the farmers, so that the number of vaccinations decreased and in 2014–2015, almost the entire costs consisted of the expenditures for monitoring and surveillance. Since 2015, financial incentives were used in the south-west of Germany (mainly Baden-Wuerttemberg) to motivate farmers to participate in voluntary vaccination, so that the vaccination costs started to increase. Since 2016, again almost all investments went into vaccinating cattle (**Figure 4**).

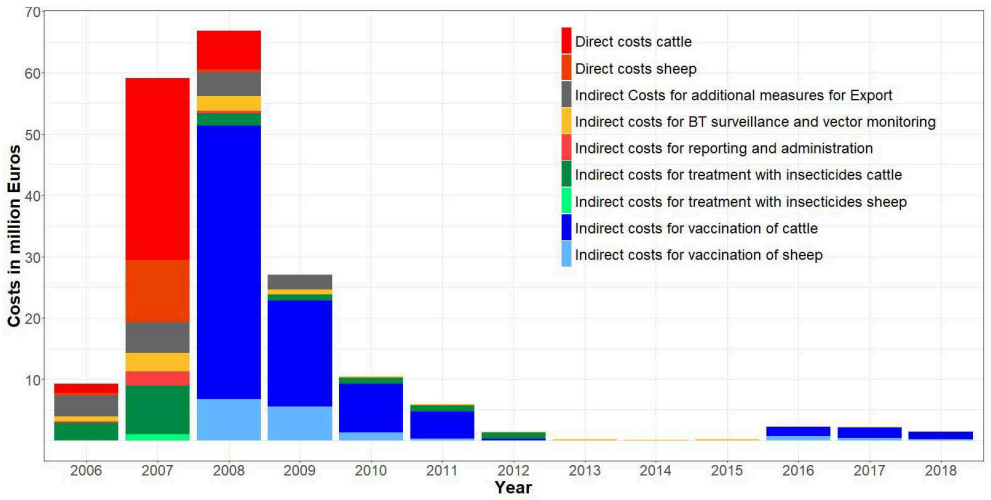


FIGURE 4 | Net total costs of BTV-8 in Germany per year from 2006 to 2018 per cost factor (mean costs by cost factors).

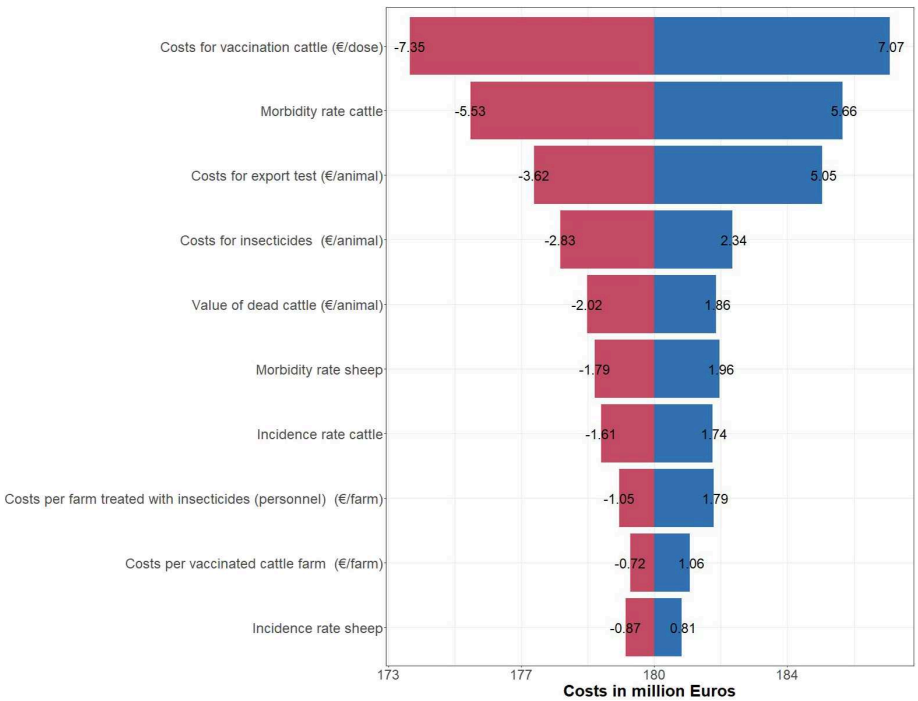


FIGURE 5 | Tornado plot for the ten most relevant cost factors of BTV-8 in Germany from 2006 to 2018. The red and the blue bars show the change in the means, when changing the single parameter.

Between 2012 and 2018, no animal trade restrictions were in place, but monitoring and surveillance were still carried out. These measures caused costs of about 150 thousand Euros in 2013–2015. In 2015, monitoring and surveillance costs increased due to the voluntary vaccination program (**Table 3** and **Figure 4**).

Sensitivity analysis showed that the proportion of infected animals that develop clinical signs and the impact on milk yield were strongly related to the veterinary treatment costs and production losses. The costs per vaccination dose had the strongest impact on the indirect costs (**Figure 5**).

DISCUSSION

When BTV-8 emerged in Germany in 2006, no validated contingency plans were available, because the disease had never before occurred in the country and came completely unexpected.

Within a short time, BTV-8 had a substantial impact on animal health in the affected livestock species and far-reaching consequences on animal trade, causing severe economic losses. According to our calculations, direct costs amounted to 40 million Euros within a single year (2007). While the direct costs were mainly borne by the farmers, the main proportion of the costs (vaccination) had to be covered by the animal disease compensation funds of the German federal states, and were co-financed by the European Union in 2008 and 2009.

To capture the full impact of BTV-8, we chose a study period of 13 years, starting in 2006, when BTV-8 first emerged in Germany, and ending in 2018, 6 years after the disease-free status was re-gained. Yet, since 2018, a new voluntary vaccination campaign is in place in western and southern regions of Germany, due the evolving BT-situation in neighboring countries, in particular Belgium, France and Switzerland.

Our study supports the view that the measures taken in reaction to a disease sometimes have a greater financial impact than the production losses caused by the disease itself (8). With 74 million Euros, the compulsory vaccination campaign in 2008–2009 was the largest cost factor throughout the study period. On the other hand, it was successful in eradicating a disease that caused not only economic damage, but also substantial suffering in infected animals.

After the initial introduction of BTV-8 in 2006, the last outbreak occurred in Germany in November 2009 and the country re-gained the official BT-free status 26 months later, in February 2012. Possibly, the numbers of new infections and thus the direct costs might have decreased anyway due to immunity after natural infection, even without vaccination. Yet, we still regard compulsory vaccination as economically beneficial, as it substantially contributed to reducing virus circulation and eventually to eradicating the disease within a short period (31). Therefore, in certain epidemiological situations, financial support should be provided to farmers who are willing to vaccinate their animals as a measure to prevent clinical BT and to reduce the size of the susceptible population, which may help to limit the spread of the disease at least to some extent.

To assess the costs of BTV-8 at the animal level, we used the GMA because it is widely used by farmers and can be easily adapted to assess the economic impact both, at the animal and farm level, for any disease with clinical symptoms.

Uncertainty regarding some input parameters limited the accuracy of our estimates. A number of key input parameters for the GMA had to be derived from expert opinion. Due to the large variation in BTV-8-associated morbidity, mortality and the severity of clinical disease, as well as large price differences for milk and vaccination (performed by both private and official veterinarians) between different regions, farms and years, our estimations were highly variable. Therefore, our model might have under- or overestimated the total costs. A sensitivity analysis identified the parameters that had the highest impact on the model outcome (e.g., morbidity, disease effects on cattle and sheep; **Figure 4**).

Not only animal and farm-related variables, but also official data were sometimes difficult to parametrize. For example, the annual reports to the European Commission underwent several changes, e.g., regarding measures eligible for co-financing or maximum financial contribution limits for testing, and they do not display the data every year in the same format. For example in 2009, costs for sheep and goats could not be differentiated, so missing data had to be estimated by interpolating the values from the previous and following years.

Moreover, we could not address all factors that might have had an economic effect on BT and its control. For example, we could not include the costs or returns from investing in biosecurity due to lack of information. We could also not include the time, farmers spent on handling diseased animals, which could have been invested in other productive activities, because this time is not known and could not be reliably estimated. Moreover, BT caused significant animal welfare problems, which were not included in our study, as animal welfare does not generate profit in terms of money and animal welfare problems are difficult to translate into monetary losses. Although Germany is highly dependent on international trade in animals and animal products, losses due to trade restrictions were only included as far as the additional testing for BT is concerned that had to be performed according to EU legislation. Further effects, for example trade partners' possible reactions to the country losing its BT-free status were not included, because the respective costs were not known and could not be reliably estimated. Moreover, we did not include additional expenditures caused by trade restrictions within Germany between August 2006 and September 2007, since the 20 and 150 km restriction zones changed frequently in short intervals (days or weeks). The fact that these parameters were not included in the calculation may have led to an underestimation of costs in 2006 and 2007. Despite these limitations, our results indicate that Germany has benefited from re-gaining its BT-free status within a short time and that the country is likely to benefit again from vaccination in the event of a new BT epidemic.

The economic impact of the BTV-8 epidemic has so far been assessed for Switzerland (32) and the Netherlands (NL) (16, 33). The cost-effectiveness of various possible surveillance systems for BT in Switzerland has also been evaluated (34) and an economic evaluation of the vector monitoring programmes in Austria and Switzerland has been conducted (35). These studies had different underlying questions and aims and therefore differed in the applied methods. All approaches have their specific advantages and disadvantages, but, overall, the variability in the methods used in the studies make it difficult to compare the outcomes. In addition, the epidemiological situation in Austria, Switzerland and Germany was not comparable, especially regarding farm structure, number of affected farms and animals, and control strategies, although the legislation in the field of animal health is similar in these countries as they are EU member states (Austria and Germany) or apply EU legislation (Switzerland).

Livestock production and the epidemiological situation regarding BT in Germany is most likely comparable to the conditions in the Netherlands. The Netherlands and Germany are part of the Common Market of the European Union. Landscape structures on both sides of the border between the two countries are similar. The epidemiological situation regarding Bluetongue disease was comparable, as the BTV-8 epidemic

started in the border area between Belgium, Germany and the Netherlands in 2006 and spread rapidly in the entire region, reaching a high prevalence in the affected area of these countries (13). Farmers tend to keep more cattle per farm in the Netherland [159.9 cattle compared to 101.7 cattle in Germany (36)], while the predominant cattle breeds are similar in both countries. The main affected regions in Germany after 2006 were in the Northeast (North Rhine-Westphalia, Lower Saxony). Farm sizes and structures in this area of Germany are more similar to those in Belgium and the Netherlands as compared to the south of Germany, in particular Bavaria and Baden-Wuerttemberg.

Nonetheless, our cost estimations are relatively low as compared to the Dutch figures, especially when considering that the number of cattle in Germany is about three times higher than in the Netherlands. For 2006, we estimated a financial impact of about 9 million Euros, compared to 28–32 million Euros in the Netherlands (16, 33). In 2007, we estimated costs of about 59 million Euros, compared to 49 million Euros (33) and 164–175 million Euros (16), which is at least in part due to the differences in the applied methods. For Switzerland, the total BTV-8 disease costs including cantonal response measures have been estimated at 12.2 and 3.6 million Euros for 2008 and 2009 (32). Again, our estimations seem to be comparatively low (6.9 million and 21.000 Euros in 2008 and 2009), especially when considering that the number of cattle in Germany is about eight times higher than in Switzerland. Based on the data of the German animal compensation funds, we estimated the mortality ratio at the population level at 0.081. This is slightly lower than the estimate for mortality used by Häsler et al. (32). A study by the Scottish Government modeled the economic impact of different incursion scenarios and estimated the direct costs of BT, including reduced milk production, weight loss, mortality, veterinary treatment and testing at £ 30 million per year (37).

The main difference between our analysis and previous studies conducted for other countries consists in the estimation of the direct costs (production losses and veterinary treatment), which was relatively low (119–136 Euros per animal) in our assessment. In contrast to the study of Velthuis et al. (16), we estimated higher milk losses, but lower veterinary treatment costs and a lower morbidity. In the Dutch study, BTV-8 was assumed to decrease milk production by 5.4 kg/day for a period of 10.5 days, which resulted in a total decrease in milk production of 56 kg per infected cow (16). Another study from the Netherlands assumed the milk production in a BTV-8 infected cow to decrease by 51 to 52 kg, which corresponds to 0.3 to 0.9% of the annual production (38). In a further study, the losses for the reduction in milk production in the Netherlands were estimated to range between 3 and 94 (average 48) Euros per cow (39). A study that analyzed data of the BTV-8 epidemic in France for 2007 found that cows lost a mean of 1.2–3.4% (111–249 kg) of their total annual milk yield (40). This was higher than the losses estimated in our study, where we assumed a reduction in milk yield of about 100 kg per infected cow.

Regarding morbidity, Velthuis et al. (16) estimated 5% on the total population level and 88% for infected animals. By contrast, we estimated the number of infected (i.e., antibody-positive) cattle at about 1.29% on the national level (6.6% in the affected region) based on the results of a cross-sectional study conducted in 2007 (12). Assuming that only 5–15% of the infected animals show clinical signs, the morbidity on the population level was estimated at 0.66% (0.33–0.99%). If we had taken the Dutch morbidity figures, the direct costs for cattle in Germany would have reached 11–71 million Euros in 2006 or 13–308 million Euros in 2007. In our model, we originally used a morbidity value that is lower than the one in the model for the Netherlands. For comparison, we also recalculated our model using the morbidity rates mentioned in the publication on the Dutch data. In contrast to the Dutch studies, we did not include the costs for indoor-housing, which had the highest impact for sheep and goats in the Netherlands in 2006 (18 million Euros) (16, 33). Indoor housing was hardly practiced in Germany, although it was recommended or even required by veterinary authorities for some time early after the introduction of BTV-8 in 2006 while it was clear that indoor housing is not an effective control measure as the main Palearctic *Culicoides* vectors for BTV-8 (*C. obsoletus* and *pulicaris* complex) were found to occur also indoors (41, 42).

Regarding mortality, cow mortality ratios of 1.2, 1.3, and 1.4 for the age categories <3 days, 3 days−1 year, and >1 year, respectively have been reported in the BTV-8 epidemic in the Netherlands for 2007 (43). In our study, we assumed mortality ratios of 0.02 in adult cows and 0.03 in calves.

Regarding fertility, infected cows were five times more likely to return to service (RTS) within 56 days after the first insemination compared to non-infected cows in Dutch herds (44). The difference in time between the first and the last insemination was 101.6 days. Comparing exposed and non-exposed farms in France, RTS increased by 8–21% (45). In another study, the same authors reported an average effect of BTV-8 exposure with a 6.7% increase in RTS and 1.9% increase in short gestations (46). Regarding sheep (47), investigated an outbreak in a flock of 355 ewes in Belgium and detected an increased ratio of 15.7% in abortions. In addition, the authors found a reduction in fertility from 59–75% to 30%. Since the calving interval can directly be used in the GMA, we used this parameter (which was assumed to be prolonged by about 80 days), instead of the RTS.

For Austria, the total net costs of the BTV-8 surveillance and vaccination programmes 2005–2013 were estimated at 22.8 million Euros (48). In the same period, surveillance and vaccination cost 96.6 million Euros in Germany. This sum is relatively low, considering that the number of cattle in Germany is about six times higher than in Austria. Regarding vector monitoring, the costs for the period 2006–2010 have been estimated at 1.42 million Euros for Austria and 94,000 Euros for Switzerland (35). In Germany, vector monitoring was only performed in 2007–2008, incurring total costs of 1.2 million Euros.

In conclusion, our study shows that the BTV-8 epidemic caused high direct costs in Germany in 2007 and high indirect costs for the compulsory vaccination programme in 2008–2009. The measures taken in reaction to the emergence of BT had a greater financial impact than the production losses caused by the disease itself. It should be pointed out, however, that vaccination proved effective with regard to disease eradication within a short time, thus reducing the suffering of animals and allowing international trade without restrictions.

The tools we developed are widely applicable for analyzing the economic impact of livestock diseases at both, the animal and the national level. They were implemented in widespread software (Excel and @Risk), so that they can be easily used, also by decision makers without programming skills. The use of GMA for assessing the economic impact at the animal or farm level ensures that we speak "the same language" as farmers who are used to communicate with GMA figures, when they analyse and discuss their economic situation. @Risk (Palisade) is an Excel add-in that allows using distributions rather than fixed values, i.e., stochastic modeling. Furthermore, @Risk supports the analysis of stochastic models. It is easy to use for persons who are familiar with Excel and is also used by other groups for economic analysis [e.g., (49)], which may eventually make it easier to compare results. A disadvantage is that @Risk is not freely available and for some statistical methods, no reference is given (e.g., sensitivity analysis).

We developed an economic model to calculate the direct and indirect costs BTV-8 for the years 2006–2018 in Germany, a country where BTV-8 has been successfully eradicated in the past. The model may assist stakeholders and decision makers in the planning of future control strategies. The results of the model may be useful to decide on further preventive and control measures in the current BTV-8 epidemic in Germany. The model may also be adapted for other countries and other vector-borne diseases.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

JG, FC, and CP contributed conception and design of the study. JG and CP developed and performed the economic analyses. CP, FC, and JG wrote the manuscript.

ACKNOWLEDGMENTS

We thank Mr Rolf Heuser for compiling and providing the annual applications of the German Federal Ministry for Food and Agriculture to the European Commission for co-financing and Professor Dr. Hans-Joachim Bätza for fruitful discussions. We are grateful to the members of the Animal Health Service of the North Rhine-Westphalia for providing estimates on key parameters.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.00065/full#supplementary-material

REFERENCES

1. Rushton J, Thornton PK, Otte MJ. Methods of economic impact assessment. *Rev Sci Tech OIE.* (1999) 18:315–42. doi: 10.20506/rst.18.2.1172
2. Bennett R. The 'direct costs'of livestock disease: the development of a system of models for the analysis of 30 endemic livestock diseases in great Britain. *J Agr Econ.* (2003) 54:55–71. doi: 10.1111/j.1477-9552.2003.tb00048.x
3. Belbis G, Zientara S, Breard E, Sailleau C, Caignard G, Vitour D, et al. Bluetongue virus: from BTV-1 to BTV-27. *Adv Virus Res.* (2017) 99:161–97. doi: 10.1016/bs.aivir.2017.08.003
4. Conraths FJ, Kramer M, Freuling C, Hoffmann B, Staubach C, Gethmann J, et al. Blauzungenkrankheit in Deutschland: Klinik, Diagnostik und Epidemiologie. *Prakt Tierarzt.* (2007) 88:9–15.
5. Elbers A, Mintiens K, Staubach C, Gerbier G, Meiswinkel R, Hendrickx G, et al. Bluetongue virus serotype 8 epidemic in north-western Europe in 2006: preliminary findings. In: Newton JR, Mellor DJ, editors. *SVEPM Conference.* Helsinki: Society for Veterinary Epidemiology and Preventive Medicine (2007). p. 231–45.
6. Elbers A, Backx A, Van Der Spek A, Ekker M, Leijs P, Steijn K, et al. Beschrijvende epidemiologie van de Bluetongue virus serotype 8 uitbraken in Nederland in 2006. *Tijdschr Diergeneeskd.* (2008) 133:222–9.
7. Williamson S, Woodger N, Darpel K. Differential diagnosis of bluetongue in cattle and sheep. *Practice.* (2008) 30:242–51. doi: 10.1136/inpract.30.5.242
8. Rushton J, Lyons N. Economic impact of Bluetongue: a review of the effects on production. *Vet Ital.* (2015) 51:401–6. doi: 10.12834/VetIt.646.3183.1
9. Thiry E, Saegerman C, Guyot H, Kirten P, Losson B, Rollin F, et al. Bluetongue in northern Europe. *Vet Rec.* (2006) 159:327. doi: 10.1136/vr.159.10.327
10. Toussaint JF, Sailleau C, Mast J, Houdart P, Czaplicki G, Demeestere L, et al. Bluetongue in Belgium, 2006. *Emerg Infect Dis.* (2007) 13:614–6. doi: 10.3201/eid1304.061136
11. Conraths FJ, Gethmann JM, Staubach C, Mettenleiter TC, Beer M, Hoffmann B. Epidemiology of bluetongue virus serotype 8, Germany. *Emerg Infect Dis.* (2009) 15:433–5. doi: 10.3201/eid1503.081210
12. Gethmann J, Probst C, Fröhlich A, Ziller M, Staubach C, Conraths FJ. *Evaluation of Bluetongue surveillance in Germany. In: International Conference on Animal Health Surveillance (ICAHS).* Lyon: Epidémiol. et santé anim (2011).
13. EFSA. Report on Epidemiological analysis of the 2006 bluetongue virus serotype 8 epidemic in north-western Europe. *EFSA J.* (2007) 5:34r. doi: 10.2903/j.efsa.2007.34r
14. Elbers ARW, Backx A, Meroc E, Gerbier G, Staubach C, Hendrickx G, et al. Field observations during the bluetongue serotype 8 epidemic in 2006. I Detection of first outbreaks and clinical signs in sheep and cattle in Belgium, France and the Netherlands. *Prev Vet Med.* (2008) 87:21–30. doi: 10.1016/j.prevetmed.2008.06.004
15. Hoffmann B, Saßerath M, Thalheim S, Bunzenthal C, Strebelow G, Beer M. Bluetongue virus serotype 8 reemergence in Germany, 2007 and 2008. *Emerging Infect Dis.* (2008) 14:1421–3. doi: 10.3201/eid1409.080417
16. Velthuis AG, Saatkamp HW, Mourits MC, De Koeijer AA, Elbers AR. Financial consequences of the Dutch bluetongue serotype 8 epidemics of 2006 and 2007. *Prev Vet Med.* (2010) 93:294–304. doi: 10.1016/j.prevetmed.2009.11.007
17. Eschbaumer M, Hoffmann B, König P, Teifke JP, Gethmann J, Conraths FJ, et al. Efficacy of three inactivated vaccines against bluetongue virus serotype 8 in sheep. *Vaccine.* (2009) 27:4169–75. doi: 10.1016/j.vaccine.2009.04.056
18. Gethmann J, Hüttner K, Heyne H, Probst C, Ziller M, Beer M, et al. Comparative safety study of three inactivated BTV-8 vaccines in sheep and cattle under field conditions. *Vaccine.* (2009) 27:4118–26. doi: 10.1016/j.vaccine.2009.04.072
19. Gethmann J, Zilow V, Probst C, Elbers ARW, Conraths FJ. Why German farmers have their animals vaccinated against Bluetongue virus serotype 8: results of a questionnaire survey. *Vaccine.* (2015) 33:214–21. doi: 10.1016/j.vaccine.2014.10.025
20. Sailleau C, Bréard E, Viarouge C, Vitour D, Romey A, Garnier A, et al. Re-emergence of Bluetongue virus serotype 8 in France, 2015. *Transbound Emerg Dis.* (2017) 64:998–1000. doi: 10.1111/tbed.12453

21. O.I.E. *Immediate Notification Received on 03/11/2017*. (2017). Available online at: http://www.oie.int/wahis_2/public/wahid.php/Reviewreport/Review?page_refer=MapFullEventReport&reportid=25166 (accessed May 13, 2019).
22. Flannery J, Sanz-Bernardo B, Ashby M, Brown H, Carpenter S, Cooke L, et al. Evidence of reduced viremia, pathogenicity and vector competence in a re-emerging European strain of bluetongue virus serotype 8 in sheep. *Transbound Emerg Dis.* (2019) 66:1177–85. doi: 10.1111/tbed.13131
23. EFSA. Assessment of listing and categorisation of animal diseases within the framework of the Animal Health Law (Regulation (EU) No 2016/429): bluetongue. *EFSA J.* (2017) 15:e04957. doi: 10.2903/j.efsa.2017.4957
24. Sailleau C, Breard E, Viarouge C, Gorlier A, Leroux A, Hirchaud E, et al. Emergence of bluetongue virus serotype 4 in mainland France in November 2017. *Transbound Emerg Dis.* (2018) 65:1158–62. doi: 10.1111/tbed.12919
25. Anonymus (2019). *Blauzungenkrankheit - Impfbarometer*. Available online at: http://www.ua-bw.de/pub/beitrag.asp?subid=5&Thema_ID=8&ID=2260&lang=DE&Pdf=No (accessed July 08, 2019).
26. BMEL. *Gutachten Wege zu einer gesellschaftlich akzeptierten Nutztierhaltung*. (2015). Available online at: https://www.bmel.de/DE/Ministerium/Organisation/Beiraete/_Texte/AgrBeirGutachtenNutztierhaltung.html (accessed December 21, 2019).
27. Rushton J. *The Economics of Animal Health and Production*. Wallingford; Oxfordshire: CABI Publishing (2009).
28. Gethmann J, Probst C, Sauter-Louis C, Conraths FJ. Economic analysis of animal disease outbreaks–BSE and Bluetongue disease as examples. *Berl Munch Tierarztl Wochenschr.* (2015) 128:478–82. doi: 10.2376/0005-9366-128-478
29. Borgonovo E, Plischke E. Sensitivity analysis: a review of recent advances. *Eur J Oper Res.* (2016) 248:869–87. doi: 10.1016/j.ejor.2015.06.032
30. Pianosi F, Beven K, Freer J, Hall JW, Rougier J, Stephenson DB, et al. Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environ Model Softw*. (2016) 79:214–32. doi: 10.1016/j.envsoft.2016.02.008
31. Pioz M, Guis H, Pleydell D, Gay E, Calavas D, Durand B, et al. Did vaccination slow the spread of bluetongue in France? *PLoS ONE*. (2014) 9:e85444. doi: 10.1371/journal.pone.0085444
32. Häsler B, Howe KS, Di Labio E, Schwermer H, Stark KD. Economic evaluation of the surveillance and intervention programme for bluetongue virus serotype 8 in Switzerland. *Prev Vet Med.* (2012) 103:93–111. doi: 10.1016/j.prevetmed.2011.09.013
33. Velthuis AGJ, Saatkamp HW, Mourits MCM, Koeijer AAD, and Elbers ARW. *Kosten-baten analyse Bluetongue: schade epidemieën 2006 en 2007 en evaluatie vaccinatiestrategieën 2008*. Wageningen Universiteit (2008).
34. Hadorn DC, Racloz V, Schwermer H, Stark KD. Establishing a cost-effective national surveillance system for Bluetongue using scenario tree modelling. *Vet Res.* (2009) 40:57. doi: 10.1051/vetres/2009040
35. Pinior B, Brugger K, Kofer J, Schwermer H, Stockreiter S, Loitsch A, et al. Economic comparison of the monitoring programmes for bluetongue vectors in Austria and Switzerland. *Vet Rec*. (2015) 176:464. doi: 10.1136/vr.102979
36. Eurostat. *Table Bovine Animals by NUTS 2 Regions [ef_lsk_bovine]*. Eurostat Data Explorer (2019). Available online at: https://ec.europa.eu/eurostat/data/database (accessed January 10, 2019).
37. Anonymus. *Assessing the Economic Impact of Different Bluetongue Virus (BTV) Incursion Scenarios in Scotland. Technical report Commission Number CR/2007/56.* Scottish Government (2008). Available online at: http://www.gov.scot/Publications/2008/10/14092524/2
38. Santman-Berends IMGA, Hage JJ, Lam TJGM, Sampimon OC, Van Schaik G. The effect of bluetongue virus serotype 8 on milk production and somatic cell count in Dutch dairy cows in 2008. *J Dairy Sci.* (2011) 94:1347–54. doi: 10.3168/jds.2010-3792
39. Van Schaik G, Berends IMGA, Van Langen H, Elbers ARW, Vellema P. Seroprevalence of bluetongue serotype 8 in cattle in the Netherlands in spring 2007, and its consequences. *Vet Rec.* (2008) 163:441–4. doi: 10.1136/vr.163.15.441
40. Nusinovici S, Souty C, Seegers H, Beaudeau F, Fourichon C. Decrease in milk yield associated with exposure to bluetongue virus serotype 8 in cattle herds. *J Dairy Sci.* (2013) 96:877–88. doi: 10.3168/jds.2012-5800
41. Baldet T, Delécolle JC, Cêtre-Sossah C, Mathieu B, Meiswinkel R, Gerbier G. Indoor activity of Culicoides associated with livestock in the bluetongue virus (BTV) affected region of northern France during autumn 2006. *Prev Vet Med.* (2008) 87:84–97. doi: 10.1016/j.prevetmed.2008.06.014
42. Clausen P-H, Stephan A, Bartsch S, Jandowsky A, Hoffmann-Köhler P, Schein E, et al. Seasonal dynamics of biting midges (Diptera: Ceratopogonidae, *Culicoides* spp.) on dairy farms of Central Germany during the 2007/2008 epidemic of bluetongue. *Parasitol Res*. (2009) 105:381–6. doi: 10.1007/s00436-009-1417-x
43. Santman-Berends IM, Van Schaik G, Bartels CJ, Stegeman JA, Vellema P. Mortality attributable to bluetongue virus serotype 8 infection in Dutch dairy cows. *Vet Microbiol.* (2011) 148:183–8. doi: 10.1016/j.vetmic.2010.09.010
44. Santman-Berends IM, Hage JJ, Van Rijn PA, Stegeman JA, Van Schaik G. Bluetongue virus serotype 8 (BTV-8) infection reduces fertility of Dutch dairy cattle and is vertically transmitted to offspring. *Theriogenology.* (2010) 74:1377–84. doi: 10.1016/j.theriogenology.2010.06.008
45. Nusinovici S, Seegers H, Joly A, Beaudeau F, Fourichon C. Quantification and at-risk period of decreased fertility associated with exposure to bluetongue virus serotype 8 in naive dairy herds. *J Dairy Sci.* (2012) 95:3008–20. doi: 10.3168/jds.2011-4799
46. Nusinovici S, Seegers H, Joly A, Beaudeau F, Fourichon C. Increase in the occurrence of abortions associated with exposure to the Bluetongue virus serotype 8 in naive dairy herds. *Theriogenology.* (2012) 78:1140–51. doi: 10.1016/j.theriogenology.2012.05.010
47. Saegerman C, Bolkaerts B, Baricalla C, Raes M, Wiggers L, De Leeuw I, et al. The impact of naturally-occurring, trans-placental bluetongue virus serotype-8 infection on reproductive performance in sheep. *Vet J.* (2011) 187:72–80. doi: 10.1016/j.tvjl.2009.11.012
48. Pinior B, Lebl K, Firth C, Rubel F, Fuchs R, Stockreiter S, et al. Cost analysis of bluetongue virus serotype 8 surveillance and vaccination programmes in Austria from 2005 to 2013. *Vet J.* (2015) 206:154–60. doi: 10.1016/j.tvjl.2015.07.032
49. Hawkins A, Burdine K, Amaral-Phillips D, Costa JHC. An economic analysis of the costs associated with pre-weaning management strategies for dairy heifers. *Animals.* (2019) 9:E471. doi: 10.3390/ani9070471

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

METHODS
published: 18 February 2020
doi: 10.3389/fvets.2020.00070

Qualitative Research for One Health: From Methodological Principles to Impactful Applications

Chris Degeling[1]* *and Melanie Rock*[2]

[1] *Faculty of Social Science, Australian Centre for Health Engagement, Evidence and Values, School of Health and Society, University of Wollongong, Wollongong, NSW, Australia,* [2] *Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada*

The One Health concept has inspired a rich vein of applied research and scholarly reflection over the past decade, yet with little influence from qualitative methodologists. With this overview, we describe the underpinning assumptions, purposes, and potential pitfalls of data collection techniques and methods of data analysis in key qualitative research methodologies. Our aim is to enhance One Health collaborations involving qualitative researchers, veterinary epidemiologists, and veterinary economists. There exist several distinct traditions of qualitative research, from which we draw selectively for illustrative purposes. Notwithstanding important distinctions, we emphasize commonalities and the potential for collaborative impact. The most important commonality is a shared focus on contextualizing human behavior and experience–culturally, economically, historically, and socially. We demonstrate that in-depth attention to context can assist veterinary economists and epidemiologists in drawing lessons from the implementation of policies and programs. In other words, qualitative researchers can assist One Health teams in distilling insights from "success stories," but also from adverse events and unintended consequences. As a result, qualitative researchers can contribute to One Health research and policy discussions by formulating more accurate and contextually-relevant parameters for future quantitative studies. When performed well, qualitative methodologies can help veterinary economists and epidemiologists to develop impactful research questions, to create more accurate and contextually-relevant parameters for quantitative studies, and to develop policy recommendations and interventions that are attuned to the political and socio-cultural context of their implementation. In sketching out the properties and features of influential methodologies, we underscore the value of working with seasoned qualitative researchers to incorporate questions about "what," "how," and "why" in mixed-methods research designs.

Keywords: interdisciplinary studies, social values, research design, veterinary medicine, social sciences

OPEN ACCESS

Edited by:
Salome Dürr,
University of Bern, Switzerland

Reviewed by:
Ewaldus Wera,
Kupang State Agricultural Polytechnic, Indonesia
Barbara Wieland,
International Livestock Research Institute, Ethiopia

***Correspondence:**
Chris Degeling
degeling@uow.edu.au

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *01 December 2019*
Accepted: *28 January 2020*
Published: *18 February 2020*

Citation:
Degeling C and Rock M (2020) Qualitative Research for One Health: From Methodological Principles to Impactful Applications. Front. Vet. Sci. 7:70. doi: 10.3389/fvets.2020.00070

BACKGROUND

Over the past decade, the One Health concept has gained traction in policy, practice, and research. Thus far, economics has proved to be the social science with the most influence in One Health research and related applications (1). Economic evaluations certainly can complement epidemiological models in One Health research. Even so, social sciences besides economics could also play important roles. Here we are thinking about disciplines such as anthropology, geography, philosophy, psychology, and sociology (2). Within these social sciences, qualitative research has deep roots; and within professional fields that focus on human health, such as medicine, nursing, and public health, qualitative research has become increasingly integrated (3, 4). This article, therefore, seeks to elevate the place of qualitative research in One Health scholarship and applications.

The goal of qualitative research is the development of concepts which help us to understand human behavior and social phenomena in their natural settings, giving emphasis to the meanings, experiences, and views of those implicated and effected (3, 5). The conditions and systems that produce health in individuals and their populations have long been a focus of qualitative researchers. Public health problems (including those involving other species) are increasingly derived from interactions between human behavior and social and technical systems. Veterinary public health researchers, epidemiologists and policymakers can benefit from integrating insights from qualitative research into their understanding of animal health, not least to improve policies and interventions for the management of the incidence and risks of disease in animals and the provision of related services (6, 7). Disciplines such as sociology, geography and anthropology have made significant contribution to our capacity to understand public health problems, and they solutions that are both novel and contextually-appropriate.

We have written this article with veterinary economists and epidemiologists in mind, especially those who may wish to collaborate with qualitative researchers. Acknowledging that other disciplines relevant to One Health such as public health and the environmental sciences could benefit from greater engagement with social science methodologies, for the purposes of this special journal issue we wish to assist veterinary economists and epidemiologists with studies that combine quantitative with qualitative methodologies. Our overall aim, therefore, is to impart foundational knowledge regarding qualitative methodologies and applications. The references that we cite, in addition, could provide additional guidance for readers who wish to delve more deeply into qualitative methodologies and their potential for research regarding One Health. To begin, we call attention to the importance of interpretation and meaning in qualitative research. This emphasis leads us to compare two popular approaches, namely **framework analysis** and **grounded theory**. After outlining some of the key features of each, we distinguish methodologies from methods. Then we describe some popular methods for collecting and creating qualitative data; these include interviews, storytelling, and texts. We provide examples in which such data have been employed, synthesized and interpreted to enhance veterinary findings, so that they are better attuned to their political and socio-cultural context.

Considerable expertise and skill are required for impactful research that combines qualitative with quantitative methodologies. Integrating qualitative studies into larger mixed-methods requires robust research design and processes. Experience with and expertise in the method used is a must; insufficient expertise and practical experience can lead to a superficial and even harmful impacts. Because analysis begins at the same time as data collection in most qualitative social science research, in this article we have foregrounded a descriptive summary of different types of data analysis over the standard toolbox of qualitative data collection techniques found in methods textbook. Our purpose is to highlight that robust research processes require that the technique used for analysis should be determined as part of the initial research design—and not chosen as an afterthought once data collection is underway. Of note, researchers should design their studies from the outset with an approach to analysis in mind. By implication, veterinary economists and epidemiologists should include qualitative researchers from the outset if they intend to undertake a qualitative investigation as part of a larger multi-method study. Moreover, even as we invite our colleagues in veterinary economics and epidemiology to consider including qualitative researchers on their teams, we caution against superficiality. In our experience, qualitative researchers are best positioned to contribute when One Health projects have been designed collaboratively, with them and with communities.

FOUNDATIONS AND SHARED ASSUMPTIONS IN QUALITATIVE RESEARCH METHODOLOGIES

While both qualitative and quantitative research can be used to investigate similar topics of relevance to veterinary economics and epidemiology, each will address a different type of question. In relation to the use of an animal vaccine, for example, a quantitative study could determine the proportion and demographic characteristics of owners who give the vaccine to their animals over a set period of time. Similarly, questionnaires that try and capture people's intentions can test the relative importance of different factors thought to be important to their decisions through the statistical analysis of responses to a series of standardized closed questions [See (8, 9) for example]. However, unless participants have the opportunity to respond to open questions in their own words then the research cannot capture novel and unanticipated information. To answer questions about why some owners do not vaccinate their animals, a qualitative study could explore both anticipated and unanticipated factors that might influence people's choices, such as their beliefs and values. Anthropological or qualitative studies can answer questions such as "What are X and Y? How do X and Y vary in different circumstances, and why?" (10, 11).

Research methods are the tools or techniques used to collect or analyse data to answer questions and to achieve objectives. In contrast, research methodologies are accounts of

philosophical principles and theoretical underpinnings of the approach and methods being used in a study (12). Qualitative researchers tend to take an interpretive, naturalistic approach to the subject of the inquiry. The methods and methodologies used in qualitative research accept that there are a range of ways that we can understand and explain the world we live in, such that the focus of inquiry is on discovering and understanding the perspectives and experiences of those being researched (13). Developing this understanding moves beyond descriptive reporting. The application of theoretical knowledge, in qualitative studies, enables specific cases to be seen and described in more abstract terms (12). Systematically comparing within and between different cases can yield more detailed insights about particular contexts or practices, as well as generalizable knowledge about specific types of social phenomena that occur in different settings (14). As a general but by no means absolute rule, qualitative research methods and methodologies share some broad orientations which include: (i) a commitment to naturalism; (ii) a focus on understanding; and (iii) a flexible approach (3, 13). All of these assumptions are shared with observational methods in epidemiology (15, 16).

Naturalism is important because context influences people's behavior, and the beliefs and values that underpin their behavior. As we elaborate below, ethnographic methods are perhaps most committed to understanding how and why people behave the way they do in "real life" because the researcher spends extended amounts of time in the study setting, observing and partaking in the every-day activities, periodic events, and conversations. Having someone observe what is happening in the social world will inevitably have some impact on what is being studied. But this limitation can be managed through reflexive research practices (17, 18). Being committed to understanding how real life works in a specific place and time or setting creates valuable information because researchers who are familiar with and knowledgeable about how people live and the constraints and choices they face are more likely to grasp how social context has impacts on people's behavior.

As an example, the anthropologist Lyle Fearnley has described how, for duck and poultry farmers near Poyang Lake in Eastern China, diseases that afflict their birds are not considered not so much as a global health threat, but as a by-product of the new agricultural systems and modes of intensive farming that currently underpin the "family business." Now recognized as a key hotspot for highly pathogenic avian influenza (HPAI) emergence, poultry farmers have set up enterprises around local wetlands raising, slaughtering, and selling flocks of "farmed wild birds" such as swan geese. These "farmed" wildlife go out onto the lake each day to graze and hence they mingle with flocks of migratory birds, which number in the millions, potentially providing a bridge for the transfer of viruses between wild waterfowl and domestic poultry (19). Moreover, local duck farmers are unable to sell or transport their birds during HPAI outbreaks, leading them try to reduce feed costs and maintain the capital invested in their flock by free-grazing them on the lake. Paradoxically, these practices increase the local risk of viral transfer and spill-over (20).

Maintaining a focus on understanding the world (or the part of it we are interested in) from the perspective of the participants creates a critical stance from which social scientist can remain skeptical of widely-held assumptions and received wisdom. Qualitative research methods move beyond simply attempting to describe social phenomena to characterizing how different objects, processes or goals acquire meaning and significance. The views that are most important are those of community members, and not those of the researchers. Thus, the best line of questioning might not be: "Why don't farmers implement biosecurity protocols?" but "How do farmers make decisions about maintaining biosecurity, or not?"; and "What kind of evidence is used to make these decisions?" These open and non-judgemental lines of questioning start from the assumption that people have good reasons for their beliefs and actions, and part of the job of the researcher is to understand their priorities, their goals, and the constraints that shape their choices and behaviors. Qualitative researchers are not simply a reporter, taking notes and writing down stories. As qualitative researchers, we must also analyse those accounts, and link the empirical findings with a theoretical understanding of what happens during fieldwork.

For instance, in a detailed body of work the geographer Gareth Enticott has described how farmers in the United Kingdom rationalize expert advice and biosecurity protocols for managing bovine tuberculosis (bTB). These farmers tend to view bTB in their herds as locally contingent, down to luck and matters of spatial ordering. Twenty years ago, the Department for Environment, Food, and Rural Affairs (DEFRA) guidance for bTB prescribed spatial separation between herds on adjacent farms, limiting introductions of new stock, and keeping badgers (as a suspected reservoir host of bTB) out of barns and yards. Drawing on scientific studies, DEFRA also proposed that reducing the size of badger populations (through culling) in pastures was unrealistic—and likely to be counterproductive by causing perturbations in badger territories and populations. Evidence supporting this assessment was generally rejected by farmers as being simplistic and impractical, because it did not account for the practical experiences of farmers (21).

Instead, as Enticott describes, farmers generally maintained two ways of understanding the incidence of bTB in their herds: (1) as being predictable; and (2) as being unpredictable. Whenever bTB emerges, this paradox allows farmers to avoid blaming themselves, to claim a sense of stewardship over the countryside, and to sustain a view of themselves as still being in control (22). For farmers attempting to manage bTB in their herds, agricultural space should not be left open to negotiation. Space could not be ordered in the way the DEFRA guidelines proposed, to prevent badgers moving in and out of their farms, so the preferred solution was to limit the number of badgers inhabiting local landscapes through culling (21, 23). At the same time, ongoing controversy about the validity of the scientific evidence meant that policymakers in Wales and the UK formed partnerships with different stakeholder groups and policy actors, which resulted in different legislative responses in these jurisdictions (24).

Flexible research strategies, as demonstrated by Enticott's (21–24) work above, are often needed to complete rigorous

qualitative studies. Careful planning remains important, but the researchers need to adapt their plans as data is produced and analyzed, and as policies change. Qualitative studies may involve several interleaving and iterative research phases. Qualitative researchers might start with a literature or scoping review of textual sources, which then progress out into the field to undertake periods of participant observation, individual interviews, and information exchange and engagement with communities and policy stakeholders.

METHODOLOGICAL APPROACHES TO ANALYSIS IN QUALITATIVE HEALTH RESEARCH

As noted in the introduction, in this article we have foregrounded a descriptive summary of different types of data analysis over the standard toolbox of qualitative data collection techniques found in methods textbooks. We have taken this step to highlight that in most qualitative studies, analysis begins during data collection, so as to iteratively shape ongoing processes of inquiry. This approach permits the refinement of research questions, the development and testing of theoretical propositions, and the pursuit of new insights or lines of inquiry as a study evolves. Importantly, combining data collection and analysis also allows researchers to look for divergence; that is, examples of talk or events that run counter to initial expectations and the emerging findings. Insufficient attention to identifying and exploring the full range of alternative explanations is a common flaw in qualitative studies undertaken by poorly-trained or inexperienced investigators. Close examination of deviant or negative cases is an essential step that allows further refinement of the research questions, study design, and interpretation of the findings. Maintaining respectful skepticism or some form of continuous, critical and reflexive analysis of what the researcher is seeing and being told is an essential component of rigourous practice in qualitative research.

Interview transcripts, other texts, images and notes are the raw data—explanations and insights are developed through iterative rounds of interpretation (25). No matter the method of data collection, qualitative researchers often maintain have a separate note book or electronic document to jot down ideas, impressions, and early interpretations (26, 27). As well as opening up new lines of inquiry, as an understanding of the social phenomenon being studied begins to emerge, this practice enables qualitative researchers to remain aware of their own positions. Self-awareness is the foundation for continuous critical reflection about social positions and potential biases related to various aspects of the research process (26). Where members of a research team have in-depth knowledge and expertise in the field being studied, the process of analysis and interpretation can also draw on their own understanding of daily life and important events, along the lines of an informal ethnographic study (28, 29).

In qualitative studies, researchers set out to identify commonalities and differences within the data set, before focusing on relationships between different parts of the data. The overarching goal is to draw descriptive conclusions, explanatory conclusions, or both, clustered around emergent or pre-determined themes. Most qualitative research involves some variant of content analysis. Veterinary economists and epidemiologists will benefit from realizing that there exist some key differences between commonly used approaches in qualitative research. Each is underpinned by different sets of assumptions about how we experience the world and communicate meaning which shape the process of analysis (12). In what follows, we describe different stages of analysis commonly found in qualitative research projects. Then, for illustrative purposes, we attend to key differences between common inductive and deductive approaches, with reference to **Grounded Theory** and **Framework Analysis**.

Stage 1: Transcription and Familiarization

Data collection in qualitative research typically results in an indexed collection of textual materials (reflective notes, policy documents, news media, etc.) and/or audio-recordings of what people say in interviews or when interacting with each other in a group. Audio-recorded data is usually transcribed into textual form to aid in analysis. The first stage is for qualitative researchers to familiarize themselves with each interview using the audio recordings, transcripts and any contextual or reflective notes that were recorded at the time. This process of familiarization requires reading and re-reading the transcripts, and potentially listening the audio-recordings. Through this labor-intensive process, qualitative researchers begin to grasp which themes and concepts are most likely to be important for the purposes of interpretation, and how these preliminary insights might be relevant to the phenomena and context being studied (30, 31).

Stage 2: Coding

After they have familiarized themselves with the data, qualitative researchers typically review transcripts and generate a set of codes (a label that describes a concept or analytic category). This way of coding comprises a first step in systematically interpreting a qualitative dataset (32). The aim of coding is to classify all of the data so that segments can be compared systematically with other parts of the dataset. Qualitative researchers often work alongside one another at this stage. When two researchers independently code a handful of transcripts—annotating them with notes—that helps promote internal reliability (33). Qualitative researchers' codes may be generated **inductively**—that is, derived gradually from the data—or predetermined and applied **deductively** as a way of approaching the data, either at the beginning or part way through the analysis. Coding is thus an interactive and iterative process—sometimes described as being in dialogue with the data.

Framework Analysis has gained popularity as a **deductive approach** to qualitative health research. When undertaking a deductive analysis, the codes are usually pre-defined by an existing body of knowledge; and in framework analysis, the codebook stems from policy-relevant questions. This analytic strategy emerged in the United Kingdom to advance policy-focused research, wherein the study objectives are largely determined in advance, based on the requirements of knowledge users (34). A set of pre-set questions facilitates a process of deduction through constant comparison, by organizing

qualitative data that correspond to these questions within an organizing matrix (35). Codes can still emerge from the data to capture unexpected themes—such that this approach to analysis allows qualitative researchers to accommodate both predetermined and unanticipated objectives for the research, so long as the unanticipated objectives have relevance for the policy issue at hand. Framework Analysis was explicitly designed so that procedures and outcomes can be assessed by people other than the qualitative researchers who designed and conducted a given study. This commitment to producing data that can be *a priori* organized and readily analyzed by other researchers or policy actors limits the use of the Framework Analysis to relatively homogenous data-sets. In a study guided by Framework Analysis, all of the data must address similar topics or key issues; otherwise, researchers would not be able to apply a codebook to the dataset. For these reasons, data collection for studies that employ Framework Analysis tend to be more structured and directive than in qualitative research that is inductively oriented, aimed at advancing social theory, or both (36). Examples of Framework Analysis in veterinary research include a study of disease management decision-making in aquaculture in Bangladesh (37); the policy imperatives driving poultry vaccination in Indonesia, Thailand and Vietnam (38); and the attitudes and practices for anthrax prevention amongst farmers in Zimbabwe (39).

For more **inductive studies**, such as those based on **Grounded Theory**, qualitative researchers go backwards and forwards between recruitment of participants, fieldnotes, interview transcripts, and the processes of conceptualization, thereby trying to make sense of the data as it is generated or collected (40, 41). In the process, qualitative researchers identify a set of cases for purposes of comparison. Based on constant comparison across cases, they develop explanatory propositions, and then they examine additional cases to confirm, discard, and refine these propositions.

Grounded Theory emerged from the ethnographic tradition in American sociology, including the sociology of health, illness, and medical knowledge. Hence studies based on Grounded Theory aim mainly at generating insights about social life in the first instance, as compared with providing practical advice to guide policy and programming, as in Framework Analysis. Impactful insights about policy and programming, however, can certainly emerge from studies based on Grounded Theory. Examples of Grounded Theory in veterinary research include studies on how pastoralists make decisions about sick chronically sick animals in Cameroon (42); the effect of compensation on farmer attitudes to exotic disease reporting (43); the politics of veterinarians and feed-store vendor control of access to antibiotics in dairy farms in rural Peru (44); and accounting for variation in people's responses to the death of their animal companions (45). In addition, veterinary economists and veterinarians might wish to compare a Framework Analysis of aquaculture in Bangladesh (37) with an inductively-based analysis, also of aquaculture in Bangladesh (46), that resembles Clarke's take on Grounded Theory (41).

As illustrated by Grounded Theory, the hallmark of qualitative research based on inductive reasoning is the gradual development of a novel set of codes based directly on a specific project. The analytic process usually proceeds through constant comparison, in which each item is checked or compared with reference to the dataset as a whole to establish analytical categories (47). Initial coding should be open—such that an interpretive label or category is applied to anything in the data that might be relevant. Codes can refer to anything, for example events or institutions, values, emotions or how a participant behaved or how an interviewer felt during an interview.

Inductive coding must go beyond fine-grained description to be inclusive of variation and to generate novel insights. Qualitative researchers who rely mainly on inductive reasoning, therefore, pay close attention to the unexpected so that the developing analysis is challenged and alternative propositions are considered. Anomalies need to be explained—the key point being that categories may added at any point of the study, even during the writing and revising process, to reflect as many of the nuances in the data as possible (41, 48). When qualitative researchers become immersed in the research setting or in communities, for example, they often witness interactions or participate in events that seem surprising. The sense of surprise arises because first-hand participation in the social life of a community has revealed phenomena that depart from received wisdom, or that highlight the researchers' unconscious biases. Rather than suppress or ignore such surprises, qualitative researchers who privilege induction in their studies tend to treat surprising interactions or events as keys to understanding. Many will articulate ("crystallize") insights of relevance to the entire study out of an initial sense of surprise (49).

The anthropologist Michael Agar, therefore, argued that abduction is the analytic partner of induction in qualitative studies. By abduction, he meant the surprises that qualitative researchers absolutely must take "seriously as a signal of a difference between what you know and what you need to learn to understand and explain what just happened" (29, p. 64). Gradually, through a process of conceptual crystallization, qualitative researchers distill key characteristics of communities or influences on networks. At the same time, qualitative researchers may leverage marked similarities and differences within the dataset to generate more refined typologies, to interrogate important concepts, or explore how categories that manifest in the dataset relate to phenomena in social life (49). Consequently, qualitative researchers who emphasize inductive reasoning often need to collect or generate additional data for purposes of comparison, especially if new or unanticipated ideas come to the fore during the iterative process of fieldwork and analysis. Hence a flexible approach is important not only for qualitative researchers themselves, but also for inter-disciplinary teams that include qualitative researchers. The more experienced the qualitative researcher, the more a qualitative researcher will tend to rely on abductive and inductive reasoning, which gives rise to a greater need for flexibility than in deductively-driven qualitative methodologies, such as Framework Analysis. Nonetheless, qualitative researchers' need for flexibility can be challenging to manage in mixed-methodology research teams (3, 50).

Stage 3: Abstracting the Main Findings

The outcome from the generation and application of codes to the dataset is a taxonomy that describes and interprets the social phenomenon of research interest. In qualitative research based on deductive reasoning, analysts can usually rely on a detailed codebook derived pragmatically from policy-related debates, which they adapt and update as the project progresses. By contrast, when coding based mainly on inductive reasoning, qualitative researchers typically create all or most of the codes themselves, sometimes *de novo*, and sometimes in relation to previous scholarship of a theoretical nature about social life. Experienced researchers who specialize in qualitative methodologies vary in the extent to which their analyses revolve around applying a detailed codebook to their datasets (49).

As qualitative researchers gain momentum and analytic purchase on the research questions, whether through deduction or by combining induction with abduction, they shift their efforts toward clarification. Visual aids such as diagrams, maps and tables may assist in depicting key codes and insights to include and elaborate in subsequent analyses (33). Informed by the analytical and theoretical ideas developed during the research, qualitative researchers then invest time in refining the emergent insights and codes, typically by regrouping them as subcategories under broader categories (also known as "themes"). Then they may select key themes for further investigation. Depending on the depth and richness of the data, the lessons learned through this process may apply well-beyond the description of specific communities, settings, or networks. Qualitative researchers often refer to "transferability," as compared with generalizability, to signal their capacity to generate insights with implications for policy and planning. For example, qualitative researchers may be able to illuminate the reasons why something is happening, to the extent of predicting how different groups might respond to a situation, or identifying dysfunctional dynamics or constructive disruption within policy-relevant organizations and economic systems.

Stage 4: Interpretation of the Results

As is the case when employing quantitative research methods, the results or findings from qualitative studies need to be contextualized within the outcomes of previous qualitative and quantitative research. This can include a summary of the similarities and differences with findings of other research studies, including reflections on the relevant socio-historical and scientific context. For **deductive studies**, such as those that employ **Framework Analysis**, for example, the interpretation of the results might also include a clear set of action-guiding or policy recommendations informed by the attitudes, beliefs and values of participants (the people the issue effects) By comparison, the results of more **inductive studies** should include some discussion of the relationship between the results of the current study and established theories about social life (3). Within these efforts to describe the broader significance of the study outcomes it is also important to describe any limitations to the way the study was conducted or the generalisability of the study findings.

DATA COLLECTION METHODS AND SAMPLING CONSIDERATIONS IN QUALITATIVE RESEARCH

The types of data that can be used in qualitative research include what people say, what they do, and how they interact with each other and the world around them (3, 25). Materials and data can include observations made by the researcher and other media through which people communicate information, such as talking, images, symbols, and textual sources; for example, policy documents, scholarly and gray literatures, signs, and posters and the contents of social and news media. Key techniques for collecting data in qualitative research include: interviews, focus groups, and other group-oriented research and engagement practices such as storyboarding and deliberative methods.

A key difference between quantitative and qualitative methodologies is the way in which the study sample is conceptualized. The aim in quantitative studies is to produce a sample that is in some way statistically representative of the whole population of interest. Consequently, a probability sample is typically used. In qualitative work the sample size depends on the study aims—what you are expecting the data to do in terms of answering the research questions. Accordingly, most qualitative research uses purposive sampling. This entails explicitly selecting participants who can generate the data appropriate to meeting the research aims and objectives; while also being able to be identified as being "representative" to those who will use the research. Sometimes mixed sampling strategies (involving quotas of different types of people likely to have different perspectives) are used to generate information-rich cases where cultural variables are likely to be important analytically.

The most methodologically convincing criteria for ensuring study rigor is to sample theoretically; that is, until data saturation where no new insights are emerging and all themes, categories and variations are fully accounted for. But this level of detail is rarely achievable given the almost limitless amount of resources required to truly achieve theoretical saturation (51). So, the most practical and pragmatic answer to the question: "how many of what types of people should make up the sample for a qualitative research study?" is "however many of the range of different types of people who will be credible to users of the research."

Interviews

Interviews provide an account of an individual's experiences, thoughts and perspectives. Interviews can be with individuals, or with small groups where the focus is on individual perspectives and not interactions within the group. It is important that interviewers try to be sensitive to the language and concepts used by the person(s) being interviewed. The aim is to explore in depth the topic being discussed and not, as is often the case in poorly-designed qualitative studies conducted by inexperienced investigators, to undertake "tick-box" exercises on a predetermined list of possible factors. To ensure they are not recording an inaccurate or superficial account, interviewers must actively check they have understood the participants' meanings, rather than relying on their own assumptions. Interviews should

be conducted at the convenience of the participants. Given that the setting of an interview inevitably affects the content and what the qualitative researcher will be able to observe and to infer, qualitative interviews tend to take place in setting that is familiar to the participants, such as at their workplace, home, or a nearby public space. Overall, qualitative interviewers seek to make the participant feel as comfortable as possible.

Interview-based studies in qualitative health research do not sample seeking to attain statistical representativeness. That is because statistical representativeness is not necessary when the objective of the study is to understand social processes. Sample sizes are determined by factors such as the depth and duration of the interviews, and whether or not data saturation is being reached and no new themes on insights are emerging (52). Systematic, non-probabilistic sampling is favored because the purpose is to identify specific groups of people who live in circumstances or possess characteristics relevant to the social phenomenon of research interest. This approach to sampling can allow qualitative researchers to include a wide range of types of people who have perspectives relevant to the research, while also recruiting participants with access to relevant sources of knowledge and social networks. Alternatively, as in criterion-intensive sampling, qualitative researchers may recruit participants who represent a narrow range of characteristics and positions.

Qualitative researchers usually design an interview guide or schedule that contains a list of core questions that approach the topic of interest from different angles. Unlike the highly structured questionnaires used in quantitative interviews, in **semi-structured interviews**, most of the questions are open ended so that the participants can respond in their own words and their ideas can be explored in more detail. The type of analysis being conducted should also shape the questions. In **deductively-driven studies**—where at least some of the analytic codes are predetermined as part of the study design—the questions put to participants can be framed much more tightly around the subject or issue of research interest. By contrast, with **inductive studies**—where the analytic codes, themes and findings of interest to the researcher emerge through their interactions with the collected data—the qualitative researchers' questions need to be sufficiently broad to cover a wide range of experiences but narrow enough to allow a focus on exploring the participants perspectives and experiences (3). Independent of whether the study is **deductive** or **inductive** in orientation, semi-structured interviews need to be attuned to the participants' responses and perspectives. Therefore, the order of questions can vary, as can the content and focus of different questions as the researcher attempts to grasp what the person being interviewed means—often using the same terms and concept as the interviewee when adding supplementary questions. Also, qualitative researchers should regard silences, evasive responses, and apparent discomfort as meaningful. For instance, participants may feel reluctant to criticize local power-brokers directly or even obliquely.

In-depth interviews have less pre-set structure than semi-structured interviews. In many instances, qualitative researchers conduct in-depth interviews to explore, in detail, just one or perhaps two issues (3). Alternatively, qualitative researchers may conduct in-depth interviews that chart the life-course of participants, whether as influential actors, or as "ordinary" people whose life-stories illuminate social life (53). Ethnographic interviews (54) combine immersion in the study setting and first-hand interactions with participants (see section below). Usually ethnographic interviews take place with one participant at a time, but not always. When conducting ethnographic interviews, qualitative researchers usually go to where the person or group being studied does the activity that is of interest to the study and to talk to them in this context. The idea is to follow people in their every-day setting, while they are performing every-day activities, asking them questions about what they are doing and why (when necessary) along the way (55, 56). Observing people as they take part in activities and questioning them in the settings of daily life can draw attention to important details about the context and their behavior. Overall, an participatory approach to interviewing people can assist qualitative researchers in understanding how local meanings and practices reflect and reproduce key structures in social life (57).

Studies conducted by the authors of this paper, as well as by others, demonstrate some of the type of knowledge that can be gleaned from semi-structured and in-depth interviews. For example, we have published articles on urban dog-walking in Australia and Canada, based on interviews that all began by asking participants about how they took care of their dogs (58, 59). These articles followed on from a qualitative study of how people cared for dogs and cats with diabetes, which also involved ethnographic interviews. These interviews revealed that the people interpreted how their pets were faring by recognizing these animals as sentient selves (60, 61). The participants who lived with diabetes themselves regarded their diabetic pets as being akin to them, even as a kindred spirit in one memorable example. Focusing on a similar line of inquiry, Vanessa Ashall and Pru Hobson-West conducted 21 semi-structured interviews in the UK with people who put forward their pet animals as donors for canine and feline blood banks. Rather than being founded purely on altruism, owners' motivations included a desire to display their identification of their animal as a member of their family, while at the same time assuaging their guilt for not volunteering themselves as donors for human blood banks (62). Studies like these illustrate how knowledge of the motivations, beliefs and understandings of animal owners around specific disease conditions or types of clinical practices can enrich epidemiological work that covers the same area, providing context and potentially explaining human behavior, the drivers of demand, and the choices that lay-people make about animal care in veterinary clinical contexts.

Numerous studies involving qualitative interviews have taken place in remote regions and resource-constrained countries. For example, by conducting semi-structured and in-depth interviews with people living on the edge of the Kibale National Park in Uganda, in addition to administering questionnaires, Paige et al. showed that local residents were highly informed about a broad range of zoonotic diseases and risks pertinent to their local area; and they also highlighted new potential sources and pathways of transmission (63). Drawing on parasitological

and epidemiological data on *Taenia solinum* from humans and domestic pigs in a remote region of PDR Lao, an interview-based project led by Kevin Bardosh showed the relationships between alcohol, ancestral sacrifice, and the consumption of uncooked (cyst-infected) pork were central to village life. As a result, health communication campaigns (advocating cooking and better hygiene, for example) had limited impacts on culturally-embedded risk behaviors—highlighting that all interventions need to be adapted to cultural settings (6). These citations are examples of rigorous and impactful contributions from qualitative researchers to the control and prevention of zoonoses (64–67), or the effectiveness of responding organizations (68, 69).

Participant Observation

By describing daily life, common routines, and unusual events in a community or network, participant-observation can assist qualitative researchers in generating novel insights about social life, and such insights can be germane to veterinary economics and epidemiology. For example, Alex Nading's study of urban households in Ciduad Sandio Nicaragua showed how biosecurity and vector control practices altered social relationships, highlighting that people in particular socio-cultural and geographical contexts interpret disease and disease control measures differently (70). This study would not have been possible without in-depth preparation and first-hand knowledge of the local context. Nading lived within the community at the time, and he spoke Spanish well-enough to conduct interviews in that language and to accompany local people as they worked. For community health workers, who were women and who were charged with intervening directly with residents to improve *Ae aegypti* control, learning about mosquito ecology and the place of Dengue fever in colonial history was the spur for reflection on and political engagement with their own situation. Rather than embrace the goals of the dengue control program, they came to identify with the female mosquitoes. In concert with a lack of scientific rigor and an overdependence on already overworked and strained volunteers, perceptions of lived-connections between community workers and female mosquitos, ultimately contributed to the failure of top-down dengue control programs in this setting (71). Nading's careful ethnographic work shows how a focus on human-animal sociality in specific places can illuminate policies informed by veterinary economics and epidemiology. In this type of social inquiry, the people and environments in and around specific sites of program implementation are highlighted, thereby yielding novel insight into global health interventions (72).

When under taking participant observation the choice of settings determines the sample (who, where, and what is observed). The choice of venues for the study is therefore critical to how well the data generated will address the research objectives, and the generalisability of any findings (73). As a research practice, participant-observation has strong roots in anthropology, sociology, and geography. Reflexively accounting for the effects of presence of the researchers and their guiding assumptions is key in any study that involves participant-observation (74). To be sure, as discussed earlier in this article, qualitative researchers must seek to identify unconscious biases, and then to move beyond such biases and toward a deeper understanding than what was possible at the outset. Participant-observation by qualitative researchers can bring unconscious biases to light, which may seem paradoxical (29). In other words, participant-observation harnesses the qualitative researchers' own subjectivity as a resource, rather than trying to eliminate or minimize subjectivity as an unwelcome source of bias and an obstruction to scientific knowledge.

Working With Groups

Focus groups are a form of group interview that draws on discussions between research participants to generate data, whereby the researcher acts a discussion moderator. The method is particularly useful for exploring people's knowledge and experiences. Focus groups are especially useful in approaching the study of organizational cultures and the operation of dominant cultural norms and values. In a focus group the researcher explicitly uses interactions between participants as part of the method—a schedule of questions acts as prompts for the group discussion. The assumption is that group processes allow people to explore and clarify their own perspective through explaining themselves to the group and listening to other people's perspectives (75).

Focus group studies can consist of anything between 4 and 50 groups, depending on the aims of the project and the resources available (76–78). Most studies are small in scale and part of a larger multi-method study. To capitalize on people's shared experiences, qualitative researchers usually aim for homogeneity within each focus group such that the sample is comprised of groups of similar participants. A significant limitation of focus group methods is that group dynamics can often work to silence minority voices. That said group work can also actively facilitate the discussion of otherwise unmentionable topics or provide opportunities for otherwise disempowered groups in society to raise issues that are important to them.

As an example, a multi-disciplinary team from the *Dynamic Drivers of Disease in Africa Program* conducted a series of studies of the social and cultural determinant of the incidence in Lassa fever in Sierra Leone, Henipah virus in Ghana, RVF in Kenya, and Trypanosomiasis in Zambia and Zimbabwe. A series of focus groups were conducted to complement a household survey, social mapping exercises and in-depth qualitative interviews with individuals. Focus groups allowed gender, occupation and age specific discussions to take place. These discussions revealed otherwise hidden cultural dimensions of disease risk, capturing sources of knowledge not conventionally considered in disease risk models which enriched analyses with local insights and perspectives based on local knowledges (7). Focus groups were also by Bardosh et al. in the previously mentioned rapid study of the transmission dynamics of *T. solinum* in a remote village in PDR Lao (6). These examples of multi-method field-based studies also show that generalized assessments of disease risks are only the first step. Cost effective and targeted interventions follow from understanding who gets sick, when and where, through engaging with the affected communities.

Focus groups methods can also be used to engage with experts. Victoria Ng and Jan Sargent employed focus groups

to identify criteria for the prioritization of zoonotic disease in Canada (79). Representatives of different stakeholders such as animal health professionals, human health professionals and lay-members of the public were enrolled into separate groups which determined 59 criteria which covered the spectrum of factors related to both individual and population level disease burdens. The result highlights the difficulty in prioritizing zoonotic diseases because of the number of factors that need to be considered. Involving members of the public in the process drew attention to the narrowness and heterogeneity of expert which could limit the range of criteria considered during disease prioritization exercises.

Sometimes qualitative researchers conduct interviews with "naturally occurring" groups (for example, family members, or people who work together). Technically speaking, qualitative interviews with pre-existing groups are not focus groups (3). When interviewing members of pre-existing groups, important to be aware of how hierarchy within the group may affect the data. A farm hand, for example, could feel inhibited by the presence of a manager from the same property. Accordingly, qualitative researchers may conduct a series of one-to-one interviews rather than a group interview. Nonetheless, conducting one-to-one interviews in sequence cannot eliminate the potential for intimidation. Feelings of intimidation can diminish the quality of the interview data, but consenting to be interviewed could also pose risks to the participants. For instance, participants may worry because their herd or flock harbors a zoonotic infection, which could decrease the value and saleability of their livestock, if discovered. Hence qualitative researchers must proceed carefully when designing and conducting their studies, to minimize any potential for harm, and to balance potential benefits against any potential for harm.

Storyboarding methodologies encourages a different kind of research participation in that it enables lay-people to develop and communicate their knowledge about a specific issue using stories and non-textual media (80, 81). The approach centers on the creation and/or manipulation of visual elements (photographs, symbols, and drawings) or other materials such as plastic figurines, felt cloths, charts, and maps to develop an account of the social phenomena of interest to the researcher and participants. Alternatively, qualitative researchers may draw on techniques and processes from theater, such as role-plays (82, 83). The methodological focus is on stimulating a detailed representation of people's knowledge of how something happens, what things are valued and cared for within their communities, and their expectations as to the likely consequences of an event in a particular setting. The important feature is that the process works to centralize "story" as a key medium for sharing existing data and allow meaning-making to be directed by participants, increasing the likelihood that the results reflect their understandings (84).

To do this successfully, qualitative researchers need to be competent and confident in the use of the chosen non-textual media, be responsive to the needs of the research participants, and be prepared to be flexible in their approach (85). Hence qualitative researchers must allow those taking part to express their views, to the extent of molding the study according to their preferences and interests. However, a clear advantage of storyboard techniques is that they can break down traditional hierarchies (for example related to age differences or expertise of researcher and research subject); and they can facilitate communication by allowing the people to express their ideas or experiences in non-verbal ways, and on their own terms (86). Storyboard methods are particularly effective in bringing a focus on the social, temporal and spatial aspects of events or phenomena of interest—for example, the point of entry and likely transmission pathway of infectious disease outbreak in a remote or rural setting. By way of illustration, drawing on previous dog population surveys, preliminary qualitative interviews and a disease model (67, 87), storyboarding with communities in northern Australia allowed for the co-creation of knowledge about the potential impacts of a rabies outbreak, and to explore the feasibility and acceptability of different prevention and control strategies (88).

Deliberative methodologies involve members of the public or lay-people in a structured process to learn about, discuss and develop collective solutions to complex policy problems. Unlike approaches to social research that elicit participants' perspectives or experiences, deliberative methods revolve around a two-way exchange of information between members of the public, experts, and potentially, decision makers (89). Participants undergo a process of education about the problem under consideration, with an emphasis on promoting reason-based dialogue so they can expand their views through the consideration of factual information and the views of others. These features mean that deliberative methods can be used to provide public input to decision making around policy issues that cannot be resolved solely on the basis of technical information, but also require the consideration of public values. Deliberative engagement can also allow opportunities for members of the public to reframe public and health policy problems in terms that are important to them, and promote imaginative engagement with different policy options and potential futures (90). For example, one of the authors have run a series of Citizens'/Community juries convened in eastern Australia have involved citizens and members of affected communities in discussions about how best to manage the present and future risks of Hendra virus spill-over events in their local area (91). The outcomes indicate that members of the public are likely to strongly support ecological approaches to mitigating the risks of Hendra virus risks when informed of the relevant facts and dilemmas, but there is fundamental disagreement as to the most appropriate mechanisms to regulate land use change, and, thereby, create or better protect flying fox habitat.

Deliberative events construct a form of mini-public or interest group, such that composition of participants will determine what kind of claims of "representativeness" can be made about the verdict or outcome (89). A deliberative group comprised of people who are directly affected by the issue at hand (for example service users) will provide a different perspective (and a potentially different recommendation) to that of a group comprised of otherwise disinterested citizens who are not directly impacted by the matter under consideration. Clearly a small group of participants brought into deliberation cannot

be politically or statistically representative of a much larger and diverse population. However, it is possible to aim for diversity in recruitment processes and to minimize selection bias. Participants can be selected based on socio-demographic criteria to ensure a diversity of perspectives is represented (92). Recruitment and selection of participants into deliberative groups should be organized based on the assumption that it is unrealistic to expect wide public understanding and deliberation, but it is possible to derive a sense of what informed and deliberative publics would advise from a smaller group (93).

Using deliberative methods is demanding—both in terms of time and resources. Finally, it is important to remember that different outcomes can and will occur when different groups of people are brought together to deliberate under highly similar conditions (89). Replication of an outcome across multiple events can add strength to the arguments and reasons put forward by participants at the end of each process. Divergence of outcomes between otherwise identical deliberative groups points to an enlarged range of positions and constituencies around the issue under consideration. Rather than trying to make all publics brought into deliberation respond in the same manner and come to the same conclusions, the goal of using these methods is to create the conditions where participants can engage in informed and reasoned discussions and make decisions and recommendation to policy-makers that authentically reflect their values and preferences (94).

THE VALUE OF QUALITATIVE RESEARCH METHODS TO VETERINARY EPIDEMIOLOGY AND ECONOMICS

Interactions between human and animal species are both beneficial and a source of risks to human health. The benefits and risks of our associations and interactions with other species are social patterned, such that small changes in how health and disease are distributed both within and across species boundaries. Public health has long been aware of the ways in which changes in how humans and animals co-exist can impact agricultural productivity and amplify burdens of disease (15). At the same time, our preoccupation with human health can render important dimensions of our relations with and reliance on animal species relatively invisible (95). The risks and benefits to human health of our interactions with and reliance on other species fold in on one another in complex ways such that just focusing on animal health can miss important dimensions of the bigger picture. For example, the mass culling of poultry conducted in response to highly pathogenic avian influenza outbreaks can result in stunting in children because of the loss of this vital source of protein (96, 97).

An overly medicalized model of the risks at the interface of human and animal health leaves insufficient space for a consideration of social well-being, and how this might be mediated in terms of relations with other species. The implications are that veterinary epidemiology and economics need to consider the relations between people (and how they engender or hinder health) and collaborate with and provide opportunities for other scientists with the necessary methodological expertise to both capture and take seriously how people think about their relationships with non-human species within the broader structures of social and economic systems that tie humans and animals together. Partnering with a social scientist can be a corrective to assuming that human-animal-environment interactions are somehow just natural systems, and understandable as if they exist independently of the social world that brings them into being (15, 98).

Against this background, qualitative research methods can be valuable to veterinary epidemiology and economics and instrumental to research processes and design in a number of ways. In the first instance qualitative research can describe, define and explain phenomena or areas that are not amenable to quantitative research methods. Spending time in the research setting, getting to know people and understand their points of view and daily experiences—which constitutes a form of ethnographic inquiry—is an important phase in the early stages of a research project. Familiarization and attention to what people do and what matters to them can put the core research question into context or generate novel questions for research that can be followed up by other quantitative and qualitative methodologies. Of course, until something is defined and classified appropriately it cannot be measured. Qualitative methods can be the foundation of efforts to enumerate variations in the relationship between features of the world—especially if the definition of what is being studied, a group of dogs at-large in public spaces (which might be a pack of strays, feral dogs, or free roaming dogs that are owned) for example, is unclear or ambiguous.

The second way qualitative methods can be valuable is as a supplementary or complementary study alongside quantitative work. Qualitative studies can be part of a validation process, in which a supplementary study is undertaken using a different method and the results are then compared for convergence. Or qualitative methods can be part of a multimethod approach which examines a particular phenomenon or topic on several different levels. The latter is not simply a matter of joining two techniques, and the former is not a case of tacking one on the end of a project. Though a survey may pick up the distribution of opinions of members of the public about an issue, a series of in-depth interviews will be required to access why people believe what they do, and how these beliefs inform their opinions. Different research settings and different methods allow access to different levels of knowledge and ways of acting in the social world. Combining methods can help to build a wider picture that can highlight hidden complexities and provide otherwise important context to the study findings.

Potential Pitfalls and Misuse of Qualitative Research Methods

The ultimate goal of using qualitative methods is to produce a plausible and coherent explanation of the phenomenon under investigation. Poor quality analysis in qualitative health research is anecdotal and overly descriptive, and therefore lacks critical reflection or deep insight. All research depends on the application of some form of theory (5). Those using qualitative research methods need to be aware of the way in which different theoretical starting points can lead to different

ways of doing research—which ultimately will determine the validity and usefulness of the research outcomes (12). Rigorous data collection technique and good quality analysis requires researchers who are appropriately trained, and most importantly, experienced in the methods and methodologies they are using. The reliability of study findings derived from most qualitative methods can be judged by the rigor and appropriateness of data collection and analyses processes. Demonstrating rigor and appropriateness requires the researcher to create and maintain meticulous and detailed records of interviews, observations, document searches, and the decisions (and their justification) made in each stage of the analysis. This record of the data and methods employed in collection and analysis should be able to stand independently so that another trained researcher could analyse the same data in the same way and come to the same basic conclusions.

The reliability of the analysis of qualitative data can be enhanced by organizing an independent assessment of transcripts by additional skilled qualitative researchers and comparing agreement between the analysts. Other validation strategies sometimes used in qualitative research are to present the study findings to the participants and see if they regard them as a reasonable account of their perspectives and experiences. Having more than one analyst can also provide assurances of consistency and that individual bias is not coloring data interpretation. It is important during the analyses to thoroughly explore negative or deviant cases, and to provide a coherent explanation of how the findings relate to, but are not invalidated by these variations. Social scientist also try to "triangulate" their findings by designing data collection processes in which evidence is deliberately sought from a wide range of different, independent sources (for example, comparing oral testimony with observations of peoples' behavior and textual sources such as reports from statutory bodies or news media). Because different groups are likely to have different perspectives, study findings also need to be interpreted in light of these other sources and forms of evidence.

Finally, the most significant pitfall for veterinary epidemiologist wanting to understand human motivations and actions using qualitative research methodologies is not having a trained social scientist on the investigator team. Qualitative methods are underpinned by both methodological and theoretical frameworks—and a thorough grounding in both is essential for study outcomes and findings to move beyond anecdote and basic description and achieve rigor in interpretation and explanation. Employing a method without properly considering or developing an understanding of the methodology means that the rationale for using a particular method is absent—as is the lens through which analysis takes place. Against this background it has been repeatedly observed that qualitative studies are largely absent from One Health research as currently construed, which has implications for the policy and real-world relevance of basic science and epidemiological studies that focus on the human-animal-environment interface (1, 2, 99, 100). By implication, we should prioritize efforts to recruit, train, and retain a cadre of qualitative researchers who specialize in One Health (98, 99). When performed well, qualitative methodologies can help veterinary economists and epidemiologists to develop impactful research questions, to create more accurate and contextually-relevant parameters for quantitative studies, and to develop policy recommendations and interventions that are attuned to the political and socio-cultural context of their implementation.

AUTHOR CONTRIBUTIONS

CD wrote the original draft. MR made substantial contributions to the writing of subsequent iterations. CD and MR contributed to subsequent redrafts. Both authors approved the final submission.

FUNDING

This work was funded by NHMRC grant #APP108379 and CIHR grant #MOP-130569.

REFERENCES

1. Friese C, Nuyts N. Posthumanist critique and human health: how nonhumans (could) figure in public health research. *Crit Pub Health*. (2017) 27:303–13. doi: 10.1080/09581596.2017.1294246
2. Craddock S, Hinchliffe S. One world, one health? Social science engagements with the one health agenda. *Soc Sci Med*. (2015) 129:1–4. doi: 10.1016/j.socscimed.2014.11.016
3. Green J, Thorogood N. *Qualitative Methods for Health Research*. London: Sage (2017).
4. Mykhalovskiy E, Eakin J, Beagan B, Beausoleil N, Gibson BE, Macdonald ME, et al. Beyond bare bones: critical, theoretically engaged qualitative research in public health. *Canad J Pub Health*. (2018) 109:613–21. doi: 10.17269/s41997-018-0154-2
5. Bryman A. *Social Research Methods*. 5th ed. New York, NY: Oxford University Press (2016).
6. Bardosh K, Inthavong P, Xayaheuang S, Okello AL. Controlling parasites, understanding practices: the biosocial complexity of a one health intervention for neglected zoonotic helminths in northern Lao PDR. *Soc Sci Med*. (2014) 120:215–23. doi: 10.1016/j.socscimed.2014.09.030
7. Dzingirai V, Bett B, Bukachi S, Lawson E, Mangwanya L, Scoones I, et al. Zoonotic diseases: who gets sick, and why? Explorations from Africa. *Crit Pub Health*. (2017) 27:97–110. doi: 10.1080/09581596.2016.1187260
8. Wera E, Mourits MCM, Hogeveen H. Intention of dog owners to participate in rabies control measures in Flores Island, Indonesia. *Prev Vet Med*. (2016) 126:138–50. doi: 10.1016/j.prevetmed.2016.01.029
9. Goyen KA, Wright JD, Cunneen A, Henning J. Playing with fire–what is influencing horse owners' decisions to not vaccinate their horses against deadly Hendra virus infection? *PLoS ONE*. (2017) 12:e0180062. doi: 10.1371/journal.pone.0180062
10. Pope C, Mays N. Qualitative research: reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. *BMJ*. (1995) 311:42–5. doi: 10.1136/bmj.311.6996.42
11. Stellmach D, Beshar I, Bedford J, du Cros P, Stringer B. Anthropology in public health emergencies: what is anthropology good for? *BMJ Global Health*. (2018) 3:e000534. doi: 10.1136/bmjgh-2017-000534
12. Green J, Thorogood N. *The Role of Theory. Qualitative Methods for Health Research*. London: Sage (2017). p. 29–48.
13. Steel D, Guala F. Introduction The Philosophy of Social Science Reader. New York, NY: Routeledge (2010). p. 1–6.

14. Yin R. *Case Study Research: Design and Methods*. 3rd ed. Thousand Oaks, CA: Sage (2002).
15. Green J. 'One health, one medicine' and critical public health. *Crit Pub Health*. (2012) 22:377–81. doi: 10.1080/09581596.2012.723395
16. Detels R. *Epidemiology: The Foundation of Public Health. Oxford Textbook of Global Public Health*. Vol. 6. New York, NY: OUP (2015). p. 403–10. doi: 10.1093/med/9780199661756.003.0103
17. Lecompte M, Schensul J. *Designing and Conducting Ethnographic Research*. 2nd ed. Walnut Creek, CA: AltaMira Press (2010).
18. Charmaz K, Mitchell RG. The myth of silent authorship: self, substance, and style in ethnographic writing. *Symb Interact*. (1996) 19:285–302. doi: 10.1525/si.1996.19.4.285
19. Fearnley L. Wild goose chase: the displacement of influenza research in the fields of Poyang Lake, China. *Cult Anthropol*. (2015) 30:12–35. doi: 10.14506/ca30.1.03
20. Fearnley L. After the livestock revolution: free grazing ducks snd influenza uncertainties in South China. *Med Anthropol Theory*. (2018) 5:72–98. doi: 10.17157/mat.5.3.378
21. Enticott G. Calculating nature: the case of badgers, bovine tuberculosis and cattle. *J Rural Stud*. (2001) 17:149–64. doi: 10.1016/S0743-0167(00)00051-6
22. Enticott G, Vanclay F. Scripts, animal health and biosecurity: the moral accountability of farmers' talk about animal health risks. *Health Risk Soc*. (2011) 13:293–309. doi: 10.1080/13698575.2011.575456
23. Enticott G. The spaces of biosecurity: prescribing and negotiating solutions to bovine tuberculosis. *Environ Plann A*. (2008) 40:1568–82. doi: 10.1068/a40304
24. Enticott G, Franklin A. Biosecurity, expertise and the institutional void: the case of bovine tuberculosis. *Sociol Rural*. (2009) 49:375–93. doi: 10.1111/j.1467-9523.2009.00496.x
25. Coffey A, Atkinson P. *Varieties of Data and Varieties of Analysis. Making Sense of Qualitative Data*. Thousand Oaks, CA: Sage (1996). p. 1–24.
26. Green J, Thorogood N. *Developing Qualitative Analysis*. London: Sage (2017). p. 285–311.
27. Emerson R, Fretz R, Shaw L. *Writing Ethnographic Fieldnotes*. Chicago, IL: University of Chicago Press (1995). doi: 10.7208/chicago/9780226206851.001.0001
28. Katz J. Ethical escape routes for underground ethnographers. *Am Ethnol*. (2006) 33:499–506. doi: 10.1525/ae.2006.33.4.499
29. Agar M. An ethnography by any other name. *Forum Qual Sozialforschung*. (2006) 7:36. Available online at: http://www.qualitative-research.net/index.php/fqs/article/viewFile/177/396
30. Poland B. Transcription quality as an aspect of rigor in qualitative research. *Qual Inquiry*. (1995) 1:290–310. doi: 10.1177/107780049500100302
31. Hammersley M. Reproducing or constructing? Some questions about transcription in social research. *Qual Res*. (2010) 10:553–69. doi: 10.1177/1468794110375230
32. Coffey A, Atkinson P. *Concepts and Coding. Making Sense of Qualitative Data*. Thousand Oaks, CA: Sage (1996). p. 26–53.
33. Miles M, Huberman A. *Qualitative Data Analysis: An Expanded Source Book*. London: Sage (1994).
34. Ritchie J, Spencer L. Qualitative data analysis for applied policy research In: Burgess RG, Bryman A, editors. *Analyzing Qualitative Data*. London; New York, NY: Routledge (1994). p. xii, 232.
35. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol*. (2013) 13:117. doi: 10.1186/1471-2288-13-117
36. Green J, Thorogood N. *Framework Analysis. Qualitative Methods for Health Research*. London: Sage (2017). p. 268–72.
37. Garza M, Mohan CV, Rahman M, Wieland B, Häsler B. The role of infectious disease impact in informing decision-making for animal health management in aquaculture systems in Bangladesh. *Prev Vet Med*. (2019) 167:202–13. doi: 10.1016/j.prevetmed.2018.03.004
38. Pongcharoensuk P, Adisasmito W, Sat LM, Silkavute P, Muchlisoh L, Cong Hoat P, et al. Avian and pandemic human influenza policy in South-East Asia: the interface between economic and public health imperatives. *Health Policy Plann*. (2011) 27:374–83. doi: 10.1093/heapol/czr056
39. Chikerema SM, Matope G, Pfukenyi DM. Awareness and attitude toward zoonoses with particular reference to anthrax among cattle owners in selected rural communities of Zimbabwe. *Vector Borne Zoonotic Dis*. (2013) 13:243–9. doi: 10.1089/vbz.2011.0916
40. Charmaz K, Belgrave LL. Grounded theory. In: Ritzer G, editor. *The Blackwell Encyclopedia of Sociology*. New York, NY: Blackwell Publishing (2007). doi: 10.1002/9781405165518.wbeosg070.pub2
41. Clarke AE. *Situational Analysis: Grounded Theory After the Postmodern Turn*. Thousand Oaks, CA: Sage (2005).
42. Profitós JMH, Moritz M, Garabed RB. What to do with chronically sick animals? Pastoralists' management strategies in the far north region of Cameroon. *Pastoralism*. (2013) 3:8. doi: 10.1186/2041-7136-3-8
43. Elbers A, Gorgievski-Duijvesteijn M, Van der Velden P, Loeffen W, Zarafshani K. A socio-psychological investigation into limitations and incentives concerning reporting a clinically suspect situation aimed at improving early detection of classical swine fever outbreaks. *Vet Microbiol*. (2010) 142:108–18. doi: 10.1016/j.vetmic.2009.09.051
44. Redding LE, Barg FK, Smith G, Galligan DT, Levy MZ, Hennessy S. The role of veterinarians and feed-store vendors in the prescription and use of antibiotics on small dairy farms in rural Peru. *J Dairy Sci*. (2013) 96:7349–54. doi: 10.3168/jds.2013-7045
45. Adams CL, Bonnett BN, Meek AH. Owner response to companion animal death: development of a theory and practical implications. *Canad Vet J*. (1999) 40:33–9.
46. Hinchliffe S, Butcher A, Rahman MM. The AMR problem: demanding economies, biological margins, and co-producing alternative strategies. *Palgrave Commun*. (2018) 4:142. doi: 10.1057/s41599-018-0195-4
47. Green J, Thorogood N. *Grounded Theory Qualitative Methods for Health Research*. London: Sage (2017). p. 289–95.
48. Charmaz K, Belgrave L. Qualitative interviewing and grounded theory analysis. *The SAGE Handb Interv Res*. (2012) 2:347–65. doi: 10.4135/9781452218403.n25
49. Borkan J. Immersion/crystallization. In: Crabtree B, Miller W, editors. *Doing Qualitative Research*. London: Sage (1999). p. 179–94.
50. Linder F. Slave ethics and imagining critically applied anthropology in public health research. *Med Anthropol*. (2004) 23:329–58. doi: 10.1080/01459740490515556
51. Green J, Thorogood N. *Sample Size. Qualitative Methods for Health Research*. London: Sage (2017). p. 78–81.
52. Bryman A. *Sampling in Qualitative Research. Social Research Methods*. 5th ed. New York, NY: Oxford University Press (2016). p. 407–20.
53. Agar M. Stories, background knowledge and themes: problems in the analysis of life history narrative. *Am Ethnol*. (1980) 7:223–39. doi: 10.1525/ae.1980.7.2.02a00010
54. Spradley JP. *The Ethnographic Interview*. New York, NY: Holt, Rinehart and Winston (1979). p. vii, 247.
55. Carpiano RM. Come take a walk with me: the "Go-Along" interview as a novel method for studying the implications of place for health and well-being. *Health Place*. (2009) 15:263–72. doi: 10.1016/j.healthplace.2008.05.003
56. Garcia CM, Eisenberg ME, Frerich EA, Lechner KE, Lust K. Conducting go-along interviews to understand context and promote health. *Qual Health Res*. (2012) 22:1395–403. doi: 10.1177/1049732312452936
57. Ybema S, Yanow D, Wels H, Kamsteeg F (editors). Studying everyday organizational life. In: *Organizational Ethnography: Studying the Complexities of Everyday Life*. London: SAGE (2009). p. 1–20. doi: 10.4135/9781446278925.n1
58. Degeling C, Rock M. "It was not just a walking experience": reflections on the role of care in dog-walking Health Promotion International. *Health Promot Int*. (2013) 28:397–406. doi: 10.1093/heapro/das024
59. Degeling C, Rock M, Rogers W, Riley T. Habitus and responsible dog-ownership: reconsidering the health promotion implications of 'dog-shaped' holes in people's lives. *Crit Pub Health*. (2015) 26:1–16. doi: 10.1080/09581596.2015.1026876
60. Rock M, Babinec P. Prototypes connect human diabetes with feline and canine diabetes in the context of animalhuman bonds: an anthropological analysis. *Anthrozoos*. (2010) 23:5–20. doi: 10.2752/175303710X12627079939062

61. Rock M, Babinec P. Diabetes in people, cats and dogs: biomedicine and manifold ontologies. *Med Anthropol*. (2008) 27:324–52. doi: 10.1080/01459740802427091
62. Ashall V, Hobson-West P. 'Doing good by proxy': human-animal kinship and the 'donation' of canine blood. *Sociol Health Illness*. (2017) 39:908–22. doi: 10.1111/1467-9566.12534
63. Paige SB, Malavé C, Mbabazi E, Mayer J, Goldberg TL. Uncovering zoonoses awareness in an emerging disease 'hotspot'. *Soc Sci Med*. (2015) 129:78–86. doi: 10.1016/j.socscimed.2014.07.058
64. Bardosh KL, El Berbri I, Ducrotoy M, Bouslikhane M, Ouafaa FF, Welburn SC. Zoonotic encounters at the slaughterhouse: pathways and possibilities for the control of cystic echinococcosis in northern Morocco. *J Biosoc Sci*. (2016) 48:S92–115. doi: 10.1017/S0021932015000486
65. Majekodunmi AO, Dongkum C, Idehen C, Langs DT, Welburn SC. Participatory epidemiology of endemic diseases in West African cattle–Ethnoveterinary and bioveterinary knowledge in Fulani disease control. *One Health*. (2018) 5:46–56. doi: 10.1016/j.onehlt.2018.03.001
66. Marí Saéz A, Weiss S, Nowak K, Lapeyre V, Zimmermann F, Düx A, et al. Investigating the zoonotic origin of the West African Ebola epidemic. *EMBO Mol Med*. (2015) 7:17–23. doi: 10.15252/emmm.201404792
67. Brookes VJ, Kennedy E, Dhagapan P, Ward MP. Qualitative research to design sustainable community-Based surveillance for rabies in northern Australia and Papua new Guinea. *Front Vet Sci*. (2017) 4:19. doi: 10.3389/fvets.2017.00019
68. Jerolmack C. Who's worried about turkeys? How 'organisational silos' impede zoonotic disease surveillance. *Sociol Health Illness*. (2013) 35:200–12. doi: 10.1111/j.1467-9566.2012.01501.x
69. Chien Y-J. How did international agencies perceive the avian influenza problem? The adoption and manufacture of the 'One World, One Health' framework. *Sociol. Health Illness*. (2013) 35:213–26. doi: 10.1111/j.1467-9566.2012.01534.x
70. Nading AM. Dengue mosquitoes are single mothers: biopolitics meets ecological aesthetics in Nicaraguan community health work. *Cult Anthropol*. (2012) 27:572–96. doi: 10.1111/j.1548-1360.2012.01162.x
71. Nading AM. *Mosquito Trails: Ecology, Health, and the Politics of Entanglement*. Berkely: University of California Press (2014). doi: 10.1525/california/9780520282612.001.0001
72. Nading AM. Humans, animals, and health: from ecology to entanglement. *Environ Soc*. (2013) 4:60–78. doi: 10.3167/ares.2013.040105
73. Schensul S, Schensul J, LeCompte M. *Essential Ethnographic Methods: Observations, Interviews, and questionnaires*. Walnut Creek, CA: AltaMira Press (1999).
74. LeCompte M, Schensul J. *Analyzing and Interpreting Ethnographic Data*. Walnut Creek, CA: AltaMira Press (1999).
75. Wilkinson S. Focus Group Research. In: Silverman D, editor. *Qualitative Research: Theory, Method and Practice*. London: Sage (2004). p. 177–99.
76. Barbour R. *Doing Focus Groups*. London: Sage (2008). doi: 10.4135/9781849208956
77. Agar M, MacDonald J. Focus groups and ethnography. *Human Organ*. (1995) 54:78–86. doi: 10.17730/humo.54.1.x102372362631282
78. Lehoux P, Poland B, Daudelin G. Focus group research and "the patient's view". *Soc Sci Med*. (2006) 63:2091–104. doi: 10.1016/j.socscimed.2006.05.016
79. Ng V, Sargeant JM. A stakeholder-informed approach to the identification of criteria for the prioritization of zoonoses in Canada. *PLoS ONE*. (2012) 7:e29752. doi: 10.1371/journal.pone.0029752
80. Laycock A, Walker D, Harrison N, Brands J. *Researching Indigenous Health: A Practical Guide for Researchers*. Melbourne, VIC: The Lowitja Institute (2011).
81. Smith L. *Decolonizing Methodologies: Research and Indigenous Peoples*. London: Zed Books Ltd. (2013).
82. Rossiter K, Kontos P, Colantonio A, Gilbert J, Gray J, Keightley M. Staging data: theatre as a tool for analysis and knowledge transfer in health research. *Soc Sci Med*. (2008) 66:130–46. doi: 10.1016/j.socscimed.2007.07.021
83. Agar M. The simulated situation: a methodological note. *Human Organ*. (1969) 28:322. doi: 10.17730/humo.28.4.q42250846kk3h166
84. Kovach M. *Indigenous Methodologies: Characteristics, Conversations, and Contexts*. Toronto: University of Toronto Press (2010).
85. Vindrola-Padros C, Martins A, Coyne I, Bryan G, Gibson F. From informed consent to dissemination: using participatory visual methods with young people with long-term conditions at different stages of research. *Glob Pub Health*. (2016) 11:636–50. doi: 10.1080/17441692.2016.1165718
86. Lupton D, Leahy D. Reimagining digital health education: reflections on the possibilities of the storyboarding method. *Health Educ J*. (2019) 78:633–46. doi: 10.1177/0017896919841413
87. Hudson EG, Brookes VJ, Ward MP. Assessing the risk of a canine rabies incursion in Northern Australia. *Front Vet Sci*. (2017) 4:141. doi: 10.3389/fvets.2017.00141
88. Degeling C, Brookes V, Lea T, Ward M. Rabies response, one health and more-than-human considerations in Indigenous communities in northern Australia. *Soc Sci Med*. (2018) 212:60–7. doi: 10.1016/j.socscimed.2018.07.006
89. Degeling C, Carter S, Rychetnik L. Which public and why deliberate? – A scoping review of public deliberation in public health and health policy research. *Soc Sci Med*. (2015) 131:114–21. doi: 10.1016/j.socscimed.2015.03.009
90. Blue G. Framing climate change for public deliberation: what role for interpretive social sciences and humanities? *J Environ Policy Plann*. (2016) 18:67–84. doi: 10.1080/1523908X.2015.1053107
91. Degeling C, Gilbert GL, Annand E, Taylor M, Walsh MG, Ward MP, et al. Managing the risk of Hendra virus spillover in Australia using ecological approaches: a report on three community juries. *PLoS ONE*. (2018) 13:e0209798. doi: 10.1371/journal.pone.0209798
92. Karpowitz CF, Raphael C. Ideals of inclusion in deliberation. *J Pub Delib*. (2016) 12:1–24. Available online at: https://www.publicdeliberation.net/jpd/vol12/iss2/art3/
93. Burgess MM. From 'trust us' to participatory governance: deliberative publics and science policy. *Pub Underst Sci*. (2014) 23:48–52. doi: 10.1177/0963662512472160
94. Degeling C, Rychetnik L, Street J, Thomas R, Carter SM. Influencing health policy through public deliberation: Lessons learned from two decades of Citizens'/community juries. *Soc Sci Med*. (2017) 179:166–71. doi: 10.1016/j.socscimed.2017.03.003
95. Wallace RG, Bergmann L, Kock R, Gilbert M, Hogerwerf L, Wallace R, et al. The dawn of structural one health: a new science tracking disease emergence along circuits of capital. *Soc Sci Med*. (2015) 129:68–77. doi: 10.1016/j.socscimed.2014.09.047
96. Iannotti L, Roy D. Nutritional impact of highly pathogenic avian influenza in Kenya. *Food Nutr Bull*. (2013) 34:299–309. doi: 10.1177/156482651303400302
97. Kavle JA, El-Zanaty F, Landry M, Galloway R. The rise in stunting in relation to avian influenza and food consumption patterns in Lower Egypt in comparison to Upper Egypt: results from 2005 and 2008 Demographic and Health Surveys. *BMC Pub Health*. (2015) 15:285. doi: 10.1186/s12889-015-1627-3
98. Rock M, Mykhalovskiy E, Schlich T. People, other animals and health knowledges: towards a research agenda. *Soc Sci Med*. (2009) 64:1970–6. doi: 10.1016/j.socscimed.2007.01.014
99. Bardosh K. One health: Science, politics and zoonotic disease in Africa. New York, NY: Routledge (2016). doi: 10.4324/9781315659749
100. Assmuth T, Chen X, Degeling C, Haahtela T, Irvine K, Keune H, et al. Integrative concepts and practices of health in transdisciplinary social ecology. *Socio Ecol Pract Res*. (2019). doi: 10.1007/s42532-019-00038-y. [Epub ahead of print].

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 26 February 2020
doi: 10.3389/fvets.2020.00073

Bayesian Network Modeling Applied to Feline Calicivirus Infection Among Cats in Switzerland

Gilles Kratzer[1], Fraser I. Lewis[2], Barbara Willi[3], Marina L. Meli[4,5], Felicitas S. Boretti[3], Regina Hofmann-Lehmann[4,5], Paul Torgerson[6], Reinhard Furrer[1,7] and Sonja Hartnack[6]*

[1] Department of Mathematics, University of Zurich, Zurich, Switzerland, [2] Independent Researcher, Utrecht, Netherlands, [3] Clinic for Small Animal Internal Medicine, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland, [4] Clinical Laboratory, Department of Clinical Diagnostics and Services, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland, [5] Center for Clinical Studies, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland, [6] Section of Epidemiology, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland, [7] Department of Computational Science, University of Zurich, Zurich, Switzerland

OPEN ACCESS

Edited by:
Andres M. Perez, University of Minnesota Twin Cities, United States

Reviewed by:
Norman Fenton, Queen Mary University of London, United Kingdom
Brianna R. Beechler, Oregon State University, United States

***Correspondence:**
Gilles Kratzer
gilles.kratzer@math.uzh.ch

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: 19 November 2019
Accepted: 28 January 2020
Published: 26 February 2020

Citation:
Kratzer G, Lewis FI, Willi B, Meli ML, Boretti FS, Hofmann-Lehmann R, Torgerson P, Furrer R and Hartnack S (2020) Bayesian Network Modeling Applied to Feline Calicivirus Infection Among Cats in Switzerland. Front. Vet. Sci. 7:73. doi: 10.3389/fvets.2020.00073

Bayesian network (BN) modeling is a rich and flexible analytical framework capable of elucidating complex veterinary epidemiological data. It is a graphical modeling technique that enables the visual presentation of multi-dimensional results while retaining statistical rigor in population-level inference. Using previously published case study data about feline calicivirus (FCV) and other respiratory pathogens in cats in Switzerland, a full BN modeling analysis is presented. The analysis shows that reducing the group size and vaccinating animals are the two actionable factors directly associated with FCV status and are primary targets to control FCV infection. The presence of gingivostomatitis and *Mycoplasma felis* is also associated with FCV status, but signs of upper respiratory tract disease (URTD) are not. FCV data is particularly well-suited to a network modeling approach, as both multiple pathogens and multiple clinical signs per pathogen are involved, along with multiple potentially interrelated risk factors. BN modeling is a holistic approach—all variables of interest may be mutually interdependent—which may help to address issues, such as confounding and collinear factors, as well as to disentangle directly vs. indirectly related variables. We introduce the BN methodology as an alternative to the classical uni- and multivariable regression approaches commonly used for risk factor analyses. We advise and guide researchers about how to use BNs as an exploratory data tool and demonstrate the limitations and practical issues. We present a step-by-step case study using FCV data along with all code necessary to reproduce our analyses in the open-source R environment. We compare and contrast the findings of the current case study using BN modeling with previous results that used classical regression techniques, and we highlight new potential insights. Finally, we discuss advanced methods, such as Bayesian model averaging, a common way of accounting for model uncertainty in a Bayesian network context.

Keywords: feline calicivirus, reproducible research, good modeling practice, graphical model, multivariable analysis, risk factor analysis, Bayesian network

1. INTRODUCTION

Risk factor analysis is often the primary goal of epidemiological studies. When the disease system under study is complex, there are likely many interdependent variables, including multiple interdependent outcome variables. Novel multivariate modeling approaches, such as Bayesian network (BN) modeling, may potentially reveal new epidemiological insights compared to classical statistical approaches (1) when applied to complex disease system data. We present an introduction and guide to BN modeling with complex epidemiological data and provide a case study analysis using animal welfare data. Animal welfare is an intrinsically multi-dimensional concept that cannot be measured directly. Comin et al. (2) included three animal-based welfare indicators: feather condition, mite infestation, and flock mortality. They considered two environment-based welfare indicators: the lightning quality of the barns (i.e., the quality of the lamps within the barns, whether the barns have windows, and whether they are automatically or manually regulated) and the air quality. A typical approach for dealing with multiple outcomes in animal welfare studies is to construct a composite score as the response variable and run a regression analysis. A disadvantage of this approach is that we may lose valuable insights by reducing the different welfare outcomes into a single dimension/outcome variable. Ideally, we want to retain all the richness of the original data. Rather than create a composite variable, we can instead keep all the original outcome variables by using a graphical modeling approach, and the particular type of graphical modeling methodology we consider here is Bayesian network modeling. With the increasing availability of data and the need to understand and explain ever more complex epidemiological systems, knowledge of how to effectively apply new multivariate methods, such as BN modeling may be increasingly relevant for veterinary epidemiologists.

Classical regression is the most popular method in epidemiology for performing risk factor analysis (see **Figure 1**). Regression analysis is a powerful, robust, and versatile statistical approach that estimates the relationship between two or more variables of interest. There are many types of regression analyses. At their core, they all examine the influence of one or more independent variables or factors on a dependent variable (also called the outcome or exposure variable) (3). The epidemiologist's expert decision about which variable is the response drives the regression. The philosophy for performing a risk factor analysis is to use a significance metric to extract relevant (influential) factors. However, this approach becomes unstable when the level of collinearity is too pronounced within the factors. In this context, collinearity means that some factors are (to a certain degree) predicted by a set of others, since the dataset contains redundant information. It is possible to identify and to remove redundancies, but the instability could remain due to inherent correlations in the system being studied. For identifying collinearity, some techniques have been proposed: e.g., changing estimated regression coefficients when a predictor variable is added or deleted, calculating the variance inflation factor (VIF), and deleting factors with large VIF. Some tests have been proposed, but no consensus exists on their usefulness. A statistically related problem arises when sub-selecting a limited number of variables, either due to redundancy within the data or due to computational limitations. A well-adapted approach to multivariate system epidemiology is the so-called Minimum Redundancy Maximum Relevance model (4). In the presence of collinearity, regression analysis is known to become unstable and to predict the effect of individual factors poorly. A high correlation among factors is common in epidemiology when studying biosecurity [e.g., (5)]. In this context, a direct association means that the set of variables will change the outcome when their values change. An indirect association is a correlation mediated by an intermediate set of variables. In epidemiology, it is important to model and identify variables that have a direct impact on the variable of interest. The directly associated variables are primary targets for intervention or for identifying the best candidate for knowledge-seeking from a modeling perspective. This process is called "structure discovery" in machine learning (6). The easy-to-interpret quantitative outputs and the holistic qualitative outputs of a typical BN model make it useful in observational analysis and a good alternative to classical methods.

Other popular graphical machine learning techniques exist, such as artificial neural networks (ANNs), regression trees, or random forest. At the outset, BN, ANN, and tree analyses look alike, as they rely on directional graphs. However, they are different approaches and should not be confused. We will not discuss the particular case of causal Bayesian networks in this paper because causal inference requires theoretical assumptions that are beyond the scope of this paper, and the methodology can become very field-specific. We refer the interested reader to Pearl and Mackenzie (7) for an extensive overview of modern causal modeling using graphical models. The general task addressed in this paper is, from an observational dataset, to find a suitable network that represents the relationships between the variables well using probabilistic methods. This paper seeks associations rather than causal links, i.e., exploratory analysis rather than confirmatory analysis.

This paper is structured as follows. A motivating example is presented in section 1.1. Section 2 gives a brief overview of the basic principles and an overview of the use of BNs in other fields. Section 3 gives a detailed presentation of the BN methodology, including a discussion of the key terms relevant to the BN modeling landscape. It also outlines some rules for good modeling. Section 4 lists the main commercial and non-commercial software implementations. Section 5 presents a case study with the FCV dataset. Finally, section 6 discusses the limits and misuse of BN models in epidemiology.

1.1. Motivating Example

Consider the fictitious example of an observational study about a particular disease in animal production. In the population, there are two breeds. The exposure status for each animal, the breed variable, and the disease status have been recorded (Possible values for exposure status are true or false and represent, for example, contact with sick animals. Possible disease status values are true or false). Based on this observed dataset, the task is to analyze the data. **Figure 2** presents the network (the

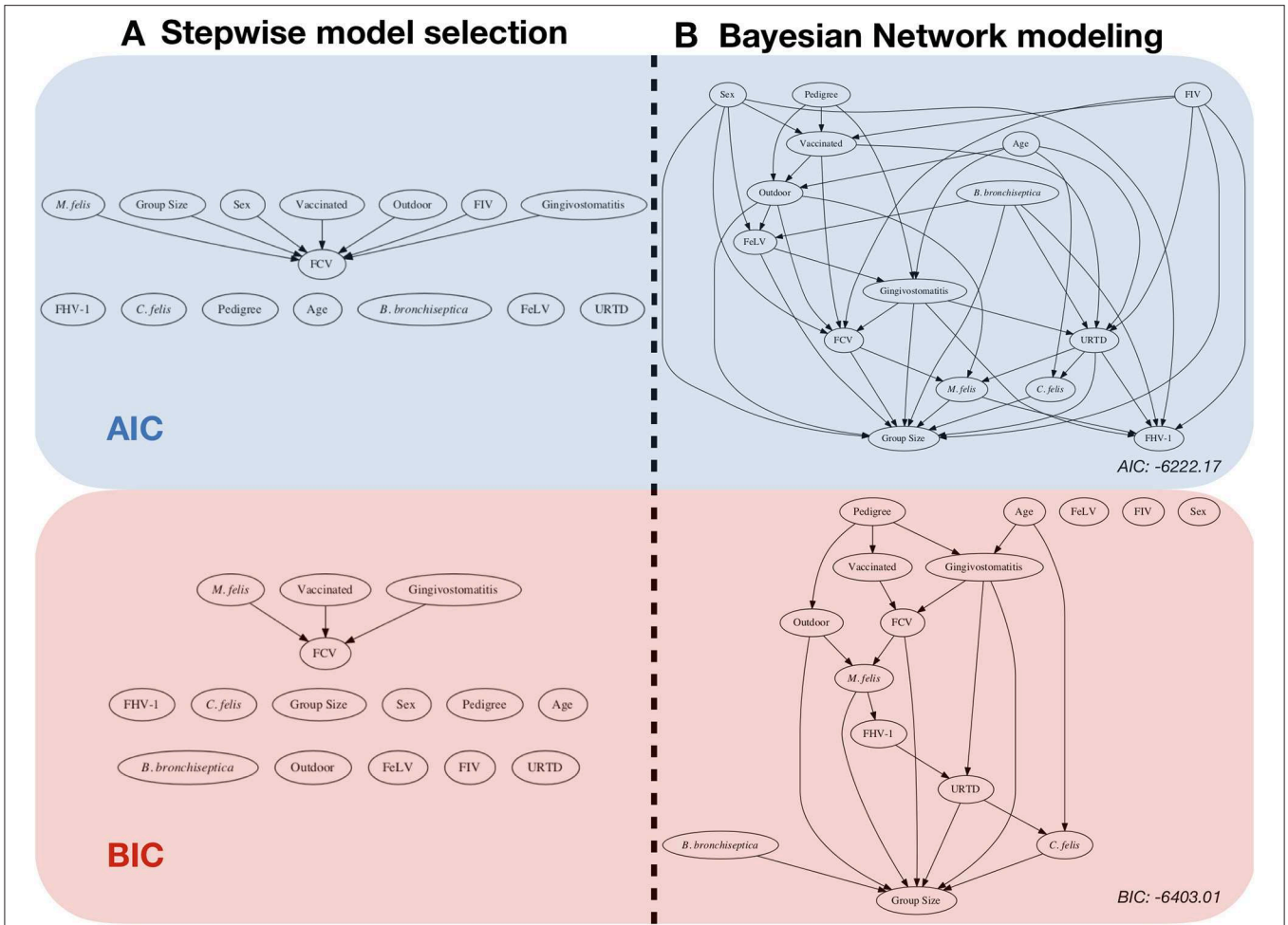


FIGURE 1 | Risk factor analysis with the FCV dataset. Comparison between: **(A)** classical stepwise multivariable model selection using two different model scores, represented by a network (backward-forward model selection algorithm), and **(B)** Bayesian network (BN) analysis with two scores and the corresponding network (using an Additive Bayesian Network algorithm). For ease of comparison, we chose to present the classical regression approach in a network format as well. The blue part was performed by using the AIC (Akaike Information criteria) score, and the red part was performed by using the BIC (Bayesian Information Criteria) score. Stepwise model selection and BN modeling using the same score gives very similar results. The BN models depict a more detailed description of the data. The objective of this paper is to describe how to obtain and to interpret such a result.

qualitative part of the model) with the so-called conditional probability tables (CPTs, the quantitative parts of the model). The CPTs are, in this discrete case and for mutually dependent variables, matrices displaying the conditional probability of a given variable with respect to the other. The classical approach would perform a regression analysis with a disease as the response and exposure and breed as factors. It would create an essentially one-dimensional BN because it would overlook the existing link between exposure and breed. We would ignore the fact that breed B is much more likely to be exposed than breed A. In epidemiology, this is a possible confounder. Indeed, breed is associated with both the dependent variable and independent variable, possibly causing a spurious association between exposure and disease status. To be classified as a confounder, breed must be causally related to exposure and disease, with no link between exposure and disease beyond the confounding effect (8). Many methods have been proposed to control for confounders in observational studies: stratification, restriction, matching, propensity score adjustment, and multiple BN regression models.

A related but different issue is the so-called effect modifier or interaction phenomenon. Interactions may arise when considering three or more variables when the effect of one variable on an outcome depends on the state of a second variable: in other words, when the effects of the two causes are not additive. An interaction can also be described as an acausal association. A typical example among humans is the interaction that exists between ultraviolet light (UV) and analogs of vitamin D (VitD) or its precursors in bone metabolism. As Lebwohl et al. (9) show, with an insufficient amount of UV light, VitD (or its precursors) will not affect bone metabolism. Symmetrically, UV light without VitD (or its precursors) will not affect bone metabolism. BN

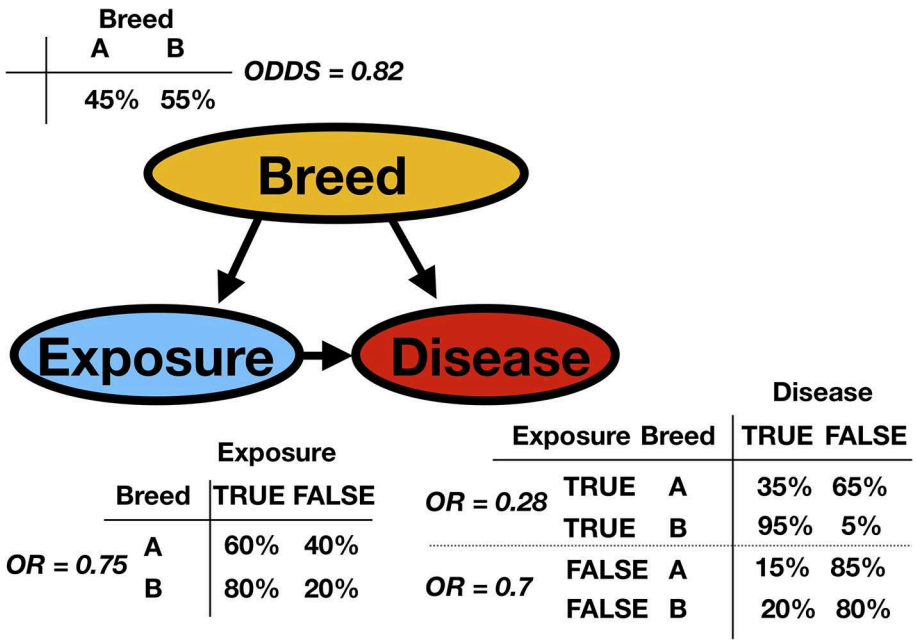


FIGURE 2 | A simple synthetic discrete Bayesian network is presented (represented as a DAG) with the conditional probability tables (CPT) and the corresponding odds or odds ratios (OR). A certain disease is studied in animal production. The observed variables are the animal's breed and the exposure to sick animals. In the DAG, *Breed* is the parent of both *Exposure* and *Disease*. *Exposure* is the parent of *Disease*, as is indicated by the arrow. Inversely, one can say that *Disease* is the child of *Exposure*. From the column or row sums of the CPTs, it is possible to extract the marginal probabilities.

modeling is conceptually attractive for performing this task in analyzing the variables within a network. From a mathematical perspective, the global model (i.e., the network and model parameter) is called the joint probability distribution.

2. BACKGROUND AND OBJECTIVES

A BN modeling approach was proposed more than 30 years ago (10). It has a track record of successful applications using real-world data in a wide variety of domains. BNs are used for modeling beliefs in social sciences (11), decision support (12), biology (13), and finance and marketing (14). More recently, this approach has been applied in veterinary epidemiology (15, 16), anti-microbial resistance (17–19), and animal welfare (2).

As BN models are used in a wide variety of research fields, they are called many different names. Here is a (non-exhaustive) list of terminology: Bayesian networks, belief networks, decision networks, probabilistic directed acyclic graphical models, recursive graphical models, naive Bayes, causal probabilistic networks, or influence diagrams (20).

Fitting BN networks to data is called *learning*. This term comes from the machine learning community and is a synonym for selecting the best network. Learning a BN from a dataset entails estimating the joint probability distribution, which encodes the global probability distribution of a multi-variable problem. When multiple variables are mutually dependent, calculating the joint probability distribution is useful as one could compute two other distributions: the marginal distribution, giving the probabilities for any variables independently of the other variables, and the conditional probability distribution, giving the probabilities for any subset of the variables conditional on particular values of the remaining variables. It is usually a two-step process involving (i) structure learning and (ii) parameter learning. This is globally called the *structure discovery* process (6). The next section presents a detailed overview of these methods. Once the joint probability distribution is estimated, it can be graphically represented using a Directed Acyclic Graph (DAG), i.e., a Bayesian Network. A BN is essentially a visual representation of a probabilistic model.

3. GENERAL METHODOLOGY

We now present the general methodology and the steps needed to fit a BN model to data. First, a short introduction to Bayesian networks is given. Then, a description of the two main learning classes of algorithm is given. Afterward, the Additive Bayesian Network (ABN) methodology is presented in detail as a special case of BN modeling. A closely related methodology is Structural Equation Modeling (SEM) (21). SEM includes different methodologies, such as confirmatory factor analysis, path analysis, partial least squares path modeling, and latent growth modeling. Although they share the same purpose, SEM and BN methodologies have significant differences (22). SEM uses a causal approach based on cause-and-effect thinking, whereas BN is based on a probabilistic approach. SEM is well-suited to latent variable modeling (i.e., variables that are not

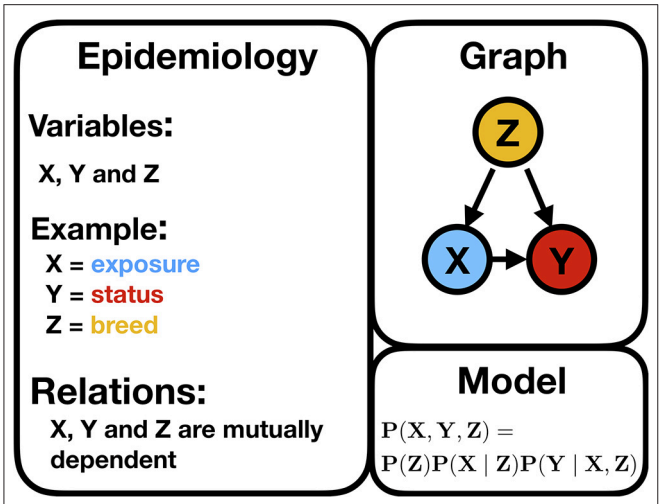


FIGURE 3 | Illustration of a typical confounding situation described in the text stating the biologically expected relationships, a DAG, and the corresponding statistical model encoding the joint probability distribution.

directly observed but are modeled from others), which is not possible in the BN methodology. This is often the primary motivation for using SEM. A BN model can take advantage of new data, whereas SEM cannot.

3.1. Bayesian Network

In a BN model context, a statistical model represents the data-generating process encoded using a graph and the parameter estimates. It can be used to describe the data, generate knowledge (i.e., understanding), or make predictions. The BN graphical representation consists of nodes, which are the random variables, and edges, which form the relationships between them. These representations often use odds ratios for discrete variables and correlation coefficients for continuous variables. The network structure should be directed and contain no cycles.

Figure 3 illustrates a typical confounder situation. In veterinary epidemiology, X could be exposure, an intervention, or a certain condition; Y is the animal status; Z is a confounder, such as sex, breed, age, or body mass index (BMI). **Figure 3** is a BN, and the following formula gives its encoded probabilistic model, indicating how to encode the joint probability distribution [i.e., $P(X, Y, Z)$] into a product of the conditional distributions

$$P(X, Y, Z) = P(Z)P(X \mid Z)P(Y \mid X, Z), \tag{1}$$

where $P(. \mid .)$ stands for the conditional probability distribution. In a discrete setting, this probability is given by the CPTs. Hence, $P(X \mid Z)$ is the conditional probability of an animal of $breed = 1$ being exposed. $P(exposure = TRUE \mid breed = 1)$. Based on the CPT, the odds ratio could be computed as the cross product of entries of the contingency table. The general formula to deduce the probabilistic model from a BN implies that the joint probability [here $P(X, Y, Z)$] factorizes as a product of the conditional probabilities of the variables, given their set of parents

$$P(\mathbf{X}) = \prod_{j=1}^{n} P(X_j \mid \mathbf{Pa}_j), \tag{2}$$

where $\mathbf{X}$ is the set of random variables (i.e., the dataset), X_j is the jth random variable, and $\mathbf{Pa}_j$ is the set of parents of the jth random variable. From a mathematical point of view, a DAG is a union of two sets: the set of nodes and the set of arrows. The network structure has a probabilistic interpretation. It encodes the factorization of the joint probability distribution of the dataset. One significant consequence of the duality between probabilistic models and network structures is that multiple different graphs can represent a given probabilistic model. As shown in **Supplementary Material**, only specific arrangements matter when selecting networks. The exact structure of a BN is not unique, and so interpretation of the effect of a variable based on arc direction, e.g., as in a typical causal statement that variable X impacts Y, is generally not valid. Therefore, on the other hand, caution is needed to avoid overinterpreting the results of a BN graph. Removing all arcs and presenting only undirected networks may potentially remove some useful information.

3.2. Model Learning

In a Bayesian setting, the model's posterior distribution given the data factorizes as the product of the structure's distribution given the data and the model parameters given the structure and the data. Then, the learning phase is formalized as:

$$P(\mathcal{M}|\mathcal{D}) = \underbrace{P(\theta_{\mathcal{M}}, \mathcal{S}|\mathcal{D})}_{\text{model learning}} = \underbrace{P(\theta_{\mathcal{M}}|\mathcal{S}, \mathcal{D})}_{\text{parameter learning}} \cdot \underbrace{P(\mathcal{S}|\mathcal{D})}_{\text{structure learning}}, \tag{3}$$

where $\mathcal{M}$ is the full model (i.e., the network structure $\mathcal{S}$ and the parameter estimates $\theta_{\mathcal{M}}$) and $\mathcal{D}$ is the dataset. One can see from Equation (3) that the two learning steps are intertwined and mutually dependent.

Learning BNs from a dataset is very complicated from the programming and statistical perspectives because the number of possible models is massive. For example, the total number of possible DAGs with 25 nodes is larger than the number of atoms in the universe (10^{80}), so the number of possible networks grows faster than the exponential function, i.e., super-exponentially. Thus, it implies the use of a smart and efficient algorithm and controlling for possible overfitting (23). In situations with limited data and numerous models, any selection or constructive method risks producing overly complicated models (i.e., a network with too many arcs) to represent the data. A key feature of the described methods is the ability to control for overfitting and so produce parsimonious models.

3.3. Structure Learning

In order to select BNs from observed data, two main approaches have been proposed: *constraint-based* and *score-based* approaches. These approaches, which are based on different statistical paradigms, are typically performed in a semi-supervised setting. Despite the intention of selecting structures in

a fully data-driven way, it is possible to guide the learning phase with external knowledge. A fully-supervised approach entails asking experts to design a network and fit it to the data. In practice, this is often not possible due to the high number of possible models that make this task highly complex. However, a semi-supervised method could be suitable, as typically, partial previous knowledge exists on specific research topics. The set of assumptions under which the learning algorithms are working are: each node in the network is a random variable (i.e., not a function of the other variables), the relationships between the random variables should be modeled with conditional independencies, every possible combination of the random variables should be plausible (even if very improbable), and the data should be derived from independent realizations of an unknown model (without temporal or spatial dependencies) (6).

3.3.1. Constraint-Based Algorithm

Constraint-based algorithms take advantage of the significant differences between colliding arrows (v-structure) and other types of structures in BN. Multiple methods exist, but one popular procedure identifies the set of mutually possible dependent variables to reduce computational complexity. It constructs the skeleton of the graph by searching for which variables are or are not related, regardless of the arc's direction. Finally, conditional independence tests are performed to detect v-structures (24). They are used as an oracle to decide on the inclusion or exclusion of an arc between two variables. Finally, based on the skeleton and the v-structure, the procedure generates (partly) directed graphs. This approach was proposed by Verma and Pearl (25). Since then, many refinements have followed. It is the methodology of choice for performing causal inference. It is also known to be more efficient with sparse networks, i.e., with a limited number of expected arcs.

Choosing the independence test framework, i.e., both the algorithm and the tests themselves, is this approach's major limitation. Another subjective user choice is how to set the significance level of the tests (classically called in statistics the α level). This choice is known to be field- and data-specific, and it influences the learned network (26). If the number of tests is substantial, precautions must be taken to avoid the problem of multiplicity, such as using Bonferroni's correction factor. Another drawback of this approach is the fact that it produces only one model.

We next present a methodology that can produce a family of plausible networks. They could be mixed to generate a more robust network. When data are scarce, it is reasonable to believe that multiple competitive models could be identified with a high level of confidence, and it would be hard to select only one.

3.3.2. Score-Based Algorithm

Bayesian network modeling can be viewed as a model selection problem. The most popular approach to BN modeling scores the candidate model in a stepwise procedure and selects the model that has the optimal score. The most popular implementation is based on an AIC (Akaike information criterion) (see **Figure 1**). The paradigm used here ensures that, if the score is well-designed, the selected model (i.e., the one with the optimal score) should represent the data well. Many scores dedicated to BN have been proposed depending on the nature of the data (for example, whether they are discrete, continuous, or a mixture of different data distributions). These scores have been designed to be penalized for model complexity because, according to Occam's razor principle, if two models for a given phenomenon exist, then the simpler should always be preferred.

To select the optimal network, i.e., the network that optimizes the network score, one needs to have a search algorithm. In contrast to the learning phase, this search algorithm only aims at finding the network with the highest possible score. Multiple algorithms have been proposed in the literature. One can perform a so-called *exact* search or use a *heuristic* approach. The *exact* search is only possible on a desktop computer for a very limited network size with a maximum of 20 nodes. Heuristic search algorithms, however, leading to an approximately optimal network, scale well with the network size (number of nodes in the network).

A score-based algorithm can learn the conditional independence between variables, so an entirely directed network could generally lead to an acausal interpretation of the arrow's direction. Thus, a score-based algorithm encodes statistical dependencies and not causal links. As Pearl (27) states:

> It seems that if conditional independence judgments are by-products of stored causal relationships, then tapping and representing those relationships directly would be a more natural and more reliable way of expressing what we know or believe about the world. This is indeed the philosophy behind causal Bayesian networks.

In a score-based perspective, arrows are important and could be displayed even if their interpretation is not fully causal. Some authors still advise that the skeleton of the network be displayed; this is also a valid approach and depends on the nature of the problem studied (28).

The major limitation of the score-based approach is the score used. Indeed, a well-designed score should minimally differentiate structures with different probabilistic models (as shown in **Figure 6**). A lot of theoretical effort has been put into deriving likelihood equivalent scores (i.e., score differentiate equivalence classes of BN), which have only been accomplished under very restrictive assumptions. For example, scores that preserve likelihood equivalence with a general mixture of data distributions do not exist. The classical workaround is to discretize the data, and then suitable scores exist. It is known in epidemiology that discretization, though common, has severe consequences and is not always advisable (29). Finally, it is interesting to note that when the number of observations is large enough, the constraint-based and score-based approaches are equivalent, and there is usually no particular reason for choosing one over the other.

3.4. Parameter Learning

Once the network structure has been selected, parameter learning can be performed locally. Only the local structure is required: the index node and the set of parent nodes. Two main approaches

exist for estimating the parameter distribution: the *maximum likelihood* and the *Bayesian* approaches. The choice of the structure learning algorithm does not influence the method for learning the parameters of the BN. Those two methods are based on two different statistical frameworks. The *maximum likelihood* assumes an unknown but fixed set of parameters for maximizing the likelihood, whereas the *Bayesian* approach treats the parameters as random and assumes a prior to them. They are computed from the posterior distribution of the parameter of the network. The main consequence is linked to the prior's choice. Indeed, the prior can help to estimate the parameters when there is not enough information within the data (30). The model parameters are interpretable as regression coefficients. Those parameters are central in model interpretation, as they give the direction of the effect and the effect size.

In the presence of missing data, more sophisticated techniques should be used to infer model parameters. In a BN context, the missing data mechanism should be ignorable (31), i.e., the data should be Missing at Random (MAR) or Missing Completely at Random (MCAR). Indeed, the MAR assumption is the minimal condition on which statistical analysis can be performed without modeling the missing data mechanism. The most popular approach for computing the value of the likelihood of the dataset with incomplete data is the Expectation-Maximization (EM) algorithm (32, 33). This is an iterative procedure for estimating the maximum *a posteriori* of a statistical model in two steps (E-step and M-step). An alternative method is variational inference, which provides a computationally cost-effective lower bound on the marginal likelihood (34). Hybrid algorithms are possible. When such solutions are not available, the usual workaround is to perform a complete case analysis, i.e., ignoring all observations containing missing information. The obvious disadvantage is the loss of existing knowledge. As an alternative, a model imputation strategy allows the researcher to still use the existing information and create data when they are missing. A good quality check is to perform both analysis and testing if they give fundamentally different results.

3.5. Additive Bayesian Network

The Additive Bayesian Network (ABN) methodology is a score-based methodology that takes advantage of a particular model parametrization. It uses the robustness and the full range of applicability of the regression framework to parameterize the network. It is used to score the candidate network and to estimate the model's parameters. The regression framework could be set in a Bayesian or a frequentist setting. The regression coefficient estimates in both Bayesian and frequentist settings are usually close to one another, but the network scores could be different, creating a very different network for each setting. Indeed, in a Bayesian setting, the so-called marginal posterior network score distribution is returned, whereas in a frequentist setting, one of the typical model selection scores is used [the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), or Minimum Distance Length (MDL)]. The term BN might be misleading, as BN models do not necessarily imply a commitment to Bayesian statistics. From a formal perspective, ABN takes advantage of the exponential family to parameterize the model and to enable the mixing of different kinds of data, such as continuous, discrete, or Poisson-distributed. A nice by-product of this parametrization is that it also allows a user to measure uncertainties of the model parameters. In a *Bayesian* setting, the credibility intervals can be computed, whereas in a *frequentist* setting, the confidence intervals can be computed. The term *additive* in ABN refers to the assumption that the effects of the variables are additive.

Figure 4 presents a scheme of the workflow used for performing an ABN analysis. A list of pre-computed scores based on atomic networks is calculated. The atomic networks are a given node with all possible combinations of parents. The list of atomic networks with their scores is called the cache of scores. Based on the cache of the pre-computed scores, a search algorithm is used to optimize the network. The search algorithm can be heuristic or exact. Based on the optimized network, the model coefficients can be fitted. ABN is thus, essentially, a graphical modeling technique that extends the usual Generalized Linear Model (GLM) to multiple dependent variables through the factorization of their joint probability distribution. In epidemiology, data are commonly generated in a setting that has an apparent grouping aspect, for example if the data are collected in different countries, counties, farms, etc. In statistics, it is known that clustering, due to the potential non-independence between data points from the same cluster, could cause over-dispersion. One of the many advantages of using the ABN framework is that it allows for adjustments for clustering within the Bayesian setting. ABN and other score-based approaches have the feature of letting the user impose external causal inputs (such as banning or retaining arcs based on previous scientific knowledge) to ensure the model's interpretability. Additionally, to considerably simplify the search space, the degenerescence between different equivalent DAGs, i.e., different networks sharing the same score, can be lifted.

ABN relies on priors at different levels. In the structure learning phase, one needs to decide on a structural prior, which encodes how likely a given structure is. In ABN, a form of prior is used that assumes that the prior probabilities for a set of parents comprising the same number of parents are all equal. It favors parents sets with either a very low or very high number of parents, which may not be appropriate. Alternatively, an uninformative prior is used where parent combinations of all cardinalities are equally likely. When using the Bayesian implementation during the model parameters learning phase, priors are used for estimation. Those priors are designed to be uninformative.

Any BN modeling approach contains approximations to make the process computationally tractable. The most common approximation is to limit the number of possible parents per node, i.e., the model complexity. Another approximation is linked to the nature of epidemiological data. Multiple types of distributions often co-exist within a dataset, and the score used should be versatile enough to handle them. From a mathematical perspective, this leads to an approximation. As previously mentioned, the chosen search algorithm could also imply some approximations. Thus, the global ABN method relies on multiple approximations, and the end-user should be aware of them. Transparently reporting them is of paramount importance.

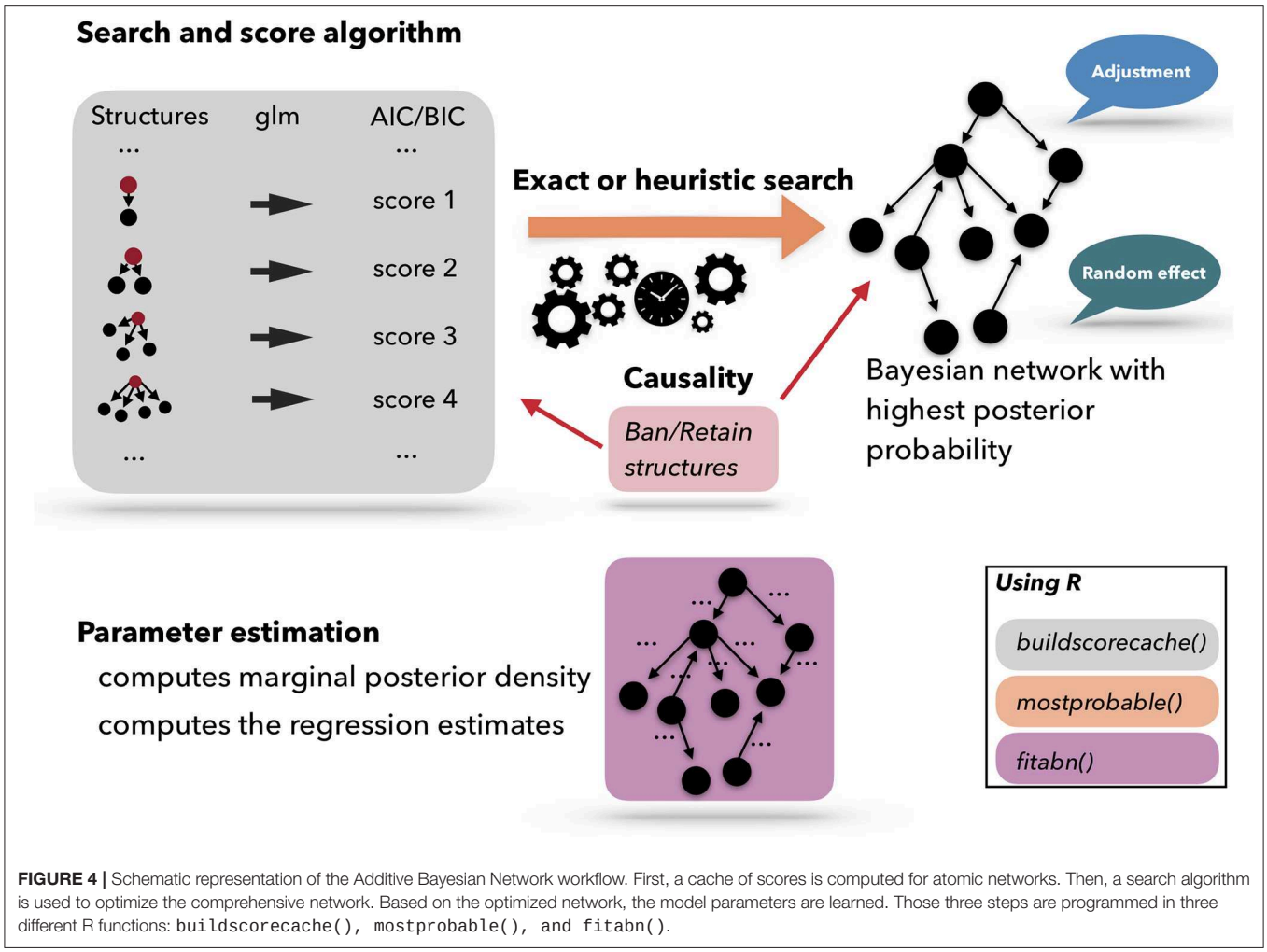


FIGURE 4 | Schematic representation of the Additive Bayesian Network workflow. First, a cache of scores is computed for atomic networks. Then, a search algorithm is used to optimize the comprehensive network. Based on the optimized network, the model parameters are learned. Those three steps are programmed in three different R functions: `buildscorecache(), mostprobable(), and fitabn()`.

4. SOFTWARE IMPLEMENTATIONS

Many commercial and non-commercial implementations for BN modeling techniques exist. Commonly commercially used BN software includes the Hugin Decision Engine produced by Hugin Expert (www.hugin.com) (35), and there is an R package to interface the Hugin Decision Engine with R: RHugin. Further examples are bayesfusion (www.bayesfusion.com), netica (www.norsys.com), and BayesiaLab (36). A popular implementation tool in MATLAB is the Bayes Net Toolbox (BNT) (37).

Within the epidemiology community, a popular open-source programming language is R (38). There are multiple R packages targeting BN modeling. The bnlearn R package contains many score-based and constraint-based algorithms, as well as multiple searching procedures (39). It is the largest and probably the most popular R package for BN modeling. When targeting causal BN inference, the pcalg R package is the most used package (40). It has a unique implementation of the PC-algorithm. The catnet R package deals with categorical data only (41). The deal R package handles both continuous and discrete variables (42). It is one of the oldest R packages for structure and parameter learning. From a more general perspective, the gRain R package is designed to perform inference in probabilistic expert systems where BNs are a special case (43). The abn R package has an implementation of a score-based system in a Bayesian and in a frequentist framework. It also has a unique implementation of an exact search algorithm and targets mixed-distributed datasets. The supported distributions are *multinomial*, *Bernoulli*, *Gaussian*, and *Poisson*. The abn R package can deal with random effects for controlling possible clustering within the data. All those packages are distributed via CRAN. Task View: gRaphical Models in R (CRAN.R-project.org/view=gR) gives a very comprehensive overview of the different computing packages available on CRAN.

5. CASE STUDY

For a case study, we focus on the Feline calicivirus (FCV) infection among cats in Switzerland. FCV is a virus that occurs worldwide in domestic cats but also in exotic felids. FCV is a highly contagious virus that is the major cause of upper respiratory tract disease or cat flu in felids. This is a disease complex caused by different viral and bacterial

TABLE 1 | Description of the factors in the FCV dataset.

Variable's name	Description
FCV	Feline calicivirus status (0/1)
FHV-1	Feline herpesvirus 1 status (0/1)
C. felis	*Chlamydia felis* status (0/1)
M. felis	*Mycoplasma felis* status (0/1)
B. bronchiseptica	*Bordetella bronchispetica* status (0/1)
FeLV	Feline leukemia virus status (0/1)
FIV	Feline immunodeficiency virus status (0/1)
Gingivostomatitis	Gingivostomatitis complex status (0/1)
URTD	Upper respiratory tract disease complex (0/1)
Vaccinated	Vaccination status (0/1)
Pedigree	Pedigree (0/1)
Outdoor	Outdoor access (0/1)
Sex	Sex and reproductive status (male, male neutered, female, female spayed)
Group size	Number of cats in the group-housing (count)
Age	Age in years (continuous)

The variable names used in R are slightly different than the ones used in the text and the figures.

pathogens, i.e., FCV, Feline Herpes Virus 1 (FHV-1), *Mycoplasma felis (M. felis)*, *Chlamydia felis (C. felis)*, and *Bordetella bronchiseptica (B. bronchoseptica)*. It can be aggravated by retrovirus infections, such as Feline Leukemia Virus (FeLV) and Feline Immunodeficiency Virus (FIV). This composite dynamic makes it very interesting for a BN modeling approach.

The data were collected between September 2012 and April 2013. Berger et al. (44) presented the original data and analysis and investigated the frequency of FCV in cats with FCV-related symptoms and in healthy cats in Switzerland. They also investigated potential protective and risk factors. The FCV dataset includes multiple viral and bacterial pathogens, retrovirus, clinical signs, and animal-related risk factors. The potential risks or protective factors are expected to be interrelated and correlated. The FCV dataset entries are described in **Table 1**. The variable sex is a composite variable between a cat's sex and reproductive status with four possible values: male, male neutered, female, female spayed.

The FCV dataset is a good candidate for a BN analysis, as complex and intertwined relations are expected among multiple recorded viruses and bacterial pathogens, animal-related variables, and environmental contributions. A major difference between this case study and the original study is that the original study design included two groups of cats: those in which FCV infection had been suspected (based on clinical signs) and healthy cats, as determined by a veterinary practitioner based on an unremarkable physical examination. The present analysis discards this study characteristic and analyzes the data as a whole observational dataset. This might hamper comparability with the original analysis so that the prevalence would not be estimable anymore.

The study enrolled 300 cats, i.e., the healthy and the FCV-suspected cats as a unique observational group. A subset of 20 of the 300 observations contain missing values. As the ABN approach requires a complete case dataset, a model imputation approach, using random forest, was used to fill in the missing data (45). Missing data are a common problem in veterinary epidemiology, and no single solution exists. However, as general advice, one can perform a complete case analysis and an imputed one. If the findings are similar, this is a good indication that there is enough information in the data to estimate an ABN model. If the findings differ significantly, then more investigations should be conducted to model the missing data. The dataset is made of 15 variables: one of them is continuous, one of them is integer-distributed, and the others are discrete.

Figure 5 presents the plots of the distributions of the individual variables. As one can see, 97 positive cases among the 300 cats are recorded. In binary logistic regressions, a popular factor of performance is the ratio between the smaller number of the two-outcome group (i.e., number of events) divided by the number of regression coefficients (excluding the intercept). In the FCV dataset, the Event Per Variable (EPV) is 97 cases divided by 7 variables (the maximum number of parents allowed), which equals 13.86 (for the *outcome: FCV*). van Smeden et al. (46) suggested that low EPV has a smaller impact than data separation or total sample size. The abn R package comes with a workaround dedicated to specifically managing low EPV and data separation: Firth's correction. The data separation problem occurs in logistic regression models when a certain combination of factors contains no observations. For example, in the FCV dataset, no record of a male cat with an FCV-positive status would imply that the sex of the cat perfectly predicts the FCV status, and the regression estimates would become numerically unstable. Firth's correction aims at producing reliable estimates in (quasi-)separated datasets. As abn tests all possible combinations of variables, the risk of data separation, especially in small datasets, is high.

5.1. Additive Bayesian Network Analysis

An ABN analysis is performed with the sequential computation of three functions: `buildscorecache()` (pre-computed scores), `mostprobable()` (a search algorithm), and `fitabn()` (parameter learning). At this stage, the user should perform multiple steps before starting an ABN analysis: loading and formatting the data, setting up the distribution of each network's nodes, deciding on possible prior knowledge, and deciding on the maximum number of parents per node (i.e., limiting the network complexity).

5.2. Loading the Data

The FCV dataset is accessible through the abn R package:

```
R> data("FCV", package = "abn")
```

5.3. Setting Up the Distributions List

The user should define a list of distributions to let abn know how to fit the data. This is similar to the family statement in the R function: `glm(..., family = binomial(link = "logit"), ...)`. One needs to create a named list that contains all the variable names and the corresponding distributions. The available distributions are *binomial, Gaussian,*

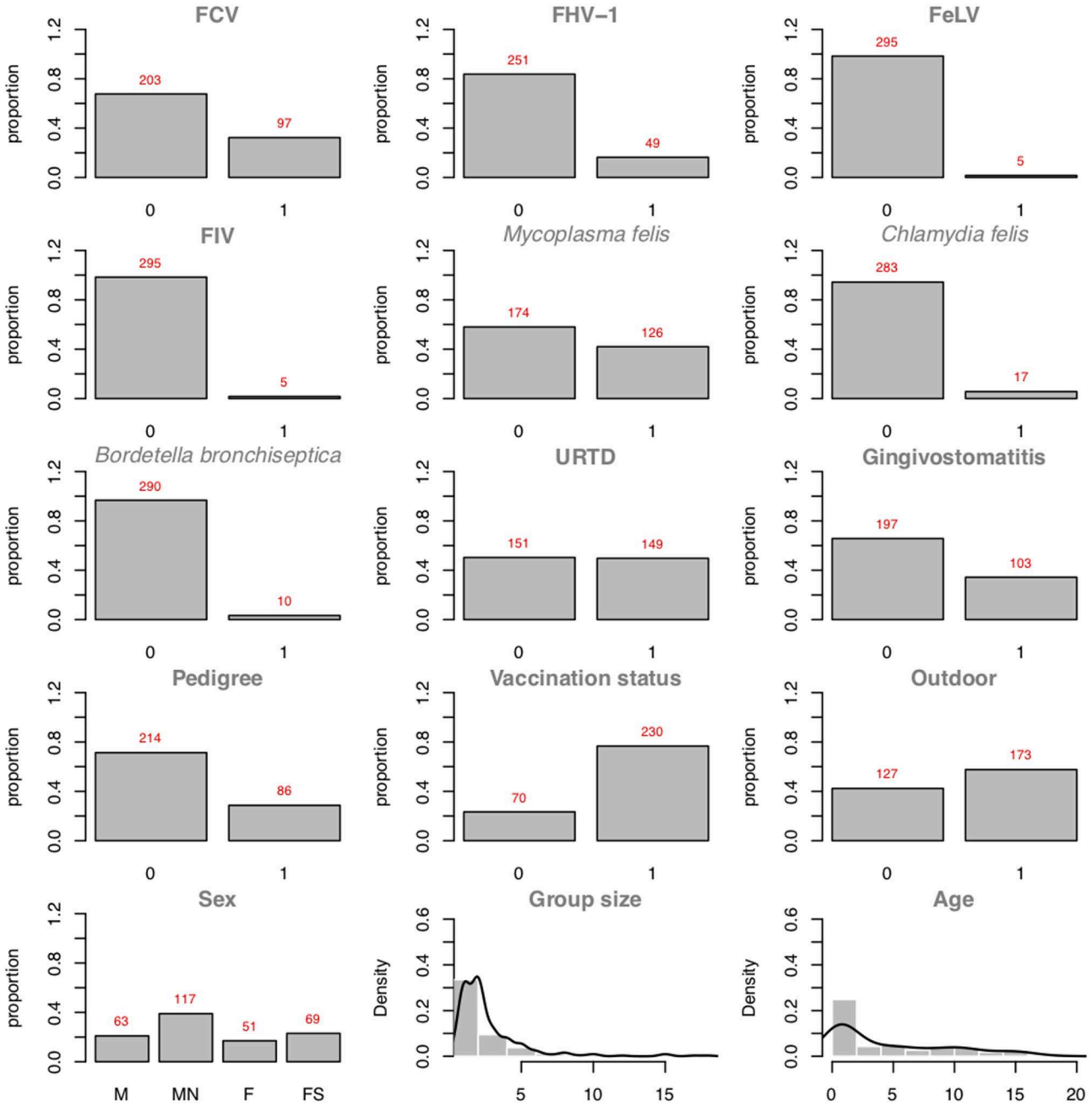


FIGURE 5 | Plots of the variable distributions in the FCV dataset. Discrete data are represented using their proportions. Continuous data are represented using histograms and their densities (*Group Size* and *Age*).

Poisson, and *multinomial*, where the last distribution is available with MLE scores only.

```
R> mydists <- list(FCV = "binomial",
+ FHV1 = "binomial",
+ FeLV = "binomial",
+ FIV = "binomial",
+ Mfelis = "binomial",
+ Cfelis = "binomial",
+ Bbronchiseptica = "binomial",
+ URTD = "binomial",
+ Gingivostomatitis = "binomial",
+ Pedigree = "binomial",
+ Vaccinated = "binomial",
+ Outdoor = "binomial",
+ Sex = "multinomial",
+ GroupSize = "poisson",
+ Age = "gaussian")
```

Binomial and multinomial data should be coerced to factors, and Gaussian and Poisson should be treated as numeric. As the list of distributions contains a *multinomial* node, the MLE method will be used in this case study.

5.4. Prior Knowledge

The prior knowledge in the abn R package can be defined by using two different means: a matrix or formula-wise statements. In the FCV dataset, the three factors *Sex, Age*, and *Pedigree* should not have a parent node. In other words, those factors cannot be influenced by any other variables within the dataset, and this prior knowledge should be transferred to abn to ensure the biological plausibility and interpretability of the final model. In practice, this is done by banning or retaining arcs within the network. By default, abn assumes no banned or retained arcs. See `?fitabn` in R about how to specify banned or retained arcs by using a formula-like syntax.

5.5. Parent Limit

To define the number of parents per node needed, one usually performs an ABN analysis in a *for loop* and increases the number of parents at each run. One computes the network score at each run and stores it. The number of parents needed is the number which leads to an unchanged network score. The code displayed below performs the so-called *parent search* for AIC, BIC and MDL scores:

```
R> aic.values <- aic.values <- mdl.values <- vector(length = 11)
R>
R> #for loop to discover the suitable network complexity
R> for (i in 1:11) {
+ max.par <- i
+ # construction of the score cache
+ mycache <- buildscorecache(data.df = mydata,
+ data.dists = dists,
+ dag.banned = ~Sex|.+Age|.+Pedigree|.,
+ max.parents = max.par, method = "mle")
+ # optimal dag with BIC
+ dag <- mostprobable(score.cache = mycache, score = "bic")
+ fabn <- fitabn(object = dag, method = "mle")
+ bic.values[i] <- fabn$bic
+ # optimal dag with AIC
+ dag <- mostprobable(score.cache = mycache, score = "aic")
+ fabn <- fitabn(object = dag, method = "mle")
+ aic.values[i] <- fabn$aic
+ # optimal dag with MDL
+ dag <- mostprobable(score.cache = mycache, score = "mdl")
+ fabn <- fitabn(object = dag, method = "mle")
+ mdl.values[i] <- fabn$mdl
+ }
```

Figure 6 displays the network score achieved in percent of the absolute maximum as a function of the maximum allowed number of parents per node for three scores (AIC, BIC, MDL). The maximum needed number of parents per node depends heavily on the chosen score. The AIC's learned network requires ten parents per node, whereas the BIC's learned network and the MDL score require only seven parents per node. This is coherent with **Figure 1**, where both stepwise model selection and BN modeling approaches with AIC select a more dense network than with BIC. Thus, choosing the score is an important modeling decision. For this case study, the BIC score will be preferred. The rationale for this subjective modeling choice is the following: since BIC is more parsimonious in terms of model complexity (considering the number of possible relationships within the network) with a limited number of observations, it is very popular with BN analysis and is closer to a Bayesian score. Based on that information, an *exact search* can be performed using the `mostprobable()` function. In the eventuality that the number of nodes would exceed 20, we would have to rely on a heuristic approach. The function `searchHillclimber()`, for example, performs multiple greedy hill-climbing searches and returns a consensus network based on a user-defined thresholding percentage. It is a good alternative when an exact search is not possible for computational reasons.

5.6. Control for Robustness and Accounting for Uncertainty

The next and final modeling step is 2-fold. It aims to control for overfitting and to account for uncertainty in the model. In statistics, and more generally in data analysis, overfitting is the production of an analysis that too closely represents the data and thus may poorly generalize findings. Overfitting produces an overly complicated model that captures unnecessary features of the studied problem. Underfitting produces a model that is too simple and thus does not capture an essential features of the studied problem. Both under- and over-fitting are limiting factors for the reliability of any analysis, but in BN modeling, the risk of overfitting is known to be high, so measures should

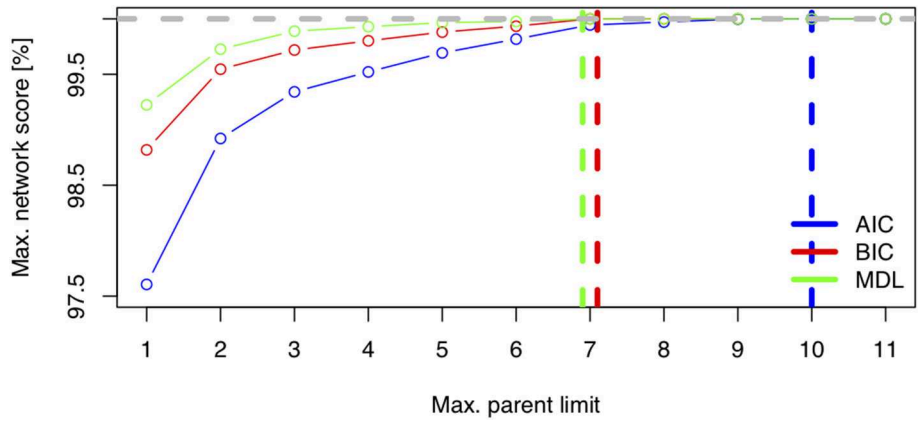


FIGURE 6 | Network score: Minimum Distance Length (MDL), Bayesian Information Criteria (BIC), and Akaike Information Criteria (AIC) as a function of network complexity, i.e., the maximum number of allowed parents per node.

be taken for controlling it. Multiple approaches have been proposed to manage the tendency of BN modeling to overfit the data. Parametric (2) and non-parametric (16) bootstrapping are very popular methods. In this case study, we use a structural Monte-Carlo Markov Chain (MCMC) sampler implemented in the mcmcabn R package (47). This approach allows us to construct MCMC samples respecting global structural priors and the chosen constraints (see **Supplementary Material** for details). The output of this modeling step estimates the probabilities of each arc's existence in the DAG. Based on those probabilities, the final DAG is pruned by removing the arcs that do not have enough support. Additionally, structural MCMC can be computationally faster for a certain class of problems.

As a by-product, the Bayesian model averaging approach helps to account for uncertainty during the modeling process. It is plausible to imagine that multiple DAGs are realistic for the data and that the limited sample size of the FCV dataset does not let us select among them objectively. Bayesian model averaging is a technique for reporting the acceptable manifold models. It has shown promising and impressive results with real-world data in closely related research fields (48–50).

From an applied perspective, so-called *structural queries* are very attractive features deducible from the MCMC sample. Structural queries are typical questions the researchers can ask the models, such as: *What is the probability that the classical signs of URTD (nasal discharge, ocular discharge, conjunctivitis, and sneezing) are NOT associated with the FCV status?* (99.7%); or *What is the probability of the gingivostomatitis complex being directly associated with the FCV status if the vaccination status is NOT?* (50%). These modeling queries are typically laborious to address with classical statistical methods, making Bayesian model averaging a very promising complementary approach to BN modeling.

5.7. Presentation of ABN Results

An ABN analysis produces qualitative results (see **Figure 7**) and quantitative results (see **Table 2**). **Figure 7** displays a pruned additive Bayesian network model constructed by pooling results across 100,000 MCMC moves. A thinning factor of 100, a burn-in phase of 17% of the total number of MCMC steps, a non-informative global network prior, and a thresholding factor of 50% were used. The original DAG was obtained using the BIC network score and seven parents per node at maximum. Three prior knowledge constraints were incorporated into the analysis: the nodes *Sex*, *Age*, and *Pedigree* cannot have any parents. The original DAG (presented in **Figure 1** in the red square on the right) has 20 arcs. The pruned one (see **Figure 7**) has 19 arcs. The square nodes are binomial, the triangular node is Poisson, the oval nodes are Gaussian, and the pentagonal node is based on multinomial distributed variables. The gray scale values of the nodes display the contribution levels of the variables (target variable, viral and bacterial pathogens, retrovirus infections, clinical signs, and animal level risk factors). The thickness of the arrow is proportional to their probability, as in classical regressions, where the effect size is almost always reported with a measure of significance. The probability of an arrow in a BN model is the counterpart of the *P*-value in a regression analysis. The percentages are reported in **Table 2** under *Support*. This table also shows the regression coefficients and their interpretation. The Confidence Intervals (CIs) are Wald-type CIs. The odds and rate ratios have a simple epidemiological interpretation. If smaller than one, a ratio has a negative effect. Inversely, if larger than one, the effect is positive. This direction of the effect is displayed in the DAG.

5.8. Interpretation of the Findings

The ABN analysis aims at studying the determinants of the FCV status. Despite using a different framework for the analysis and different datasets, the findings of the present case study are very similar to the initial results presented by Berger et al. (44). In **Figure 7**, the FCV status is directly associated with the vaccination status, the gingivostomatitis complex, the size of the housing group, and the presence of *M. felis*. The vaccination is negatively associated with the FCV status, with an odds ratio of 0.38 (i.e., the vaccinated cats are less likely to have a positive FCV status) and a supportive probability of

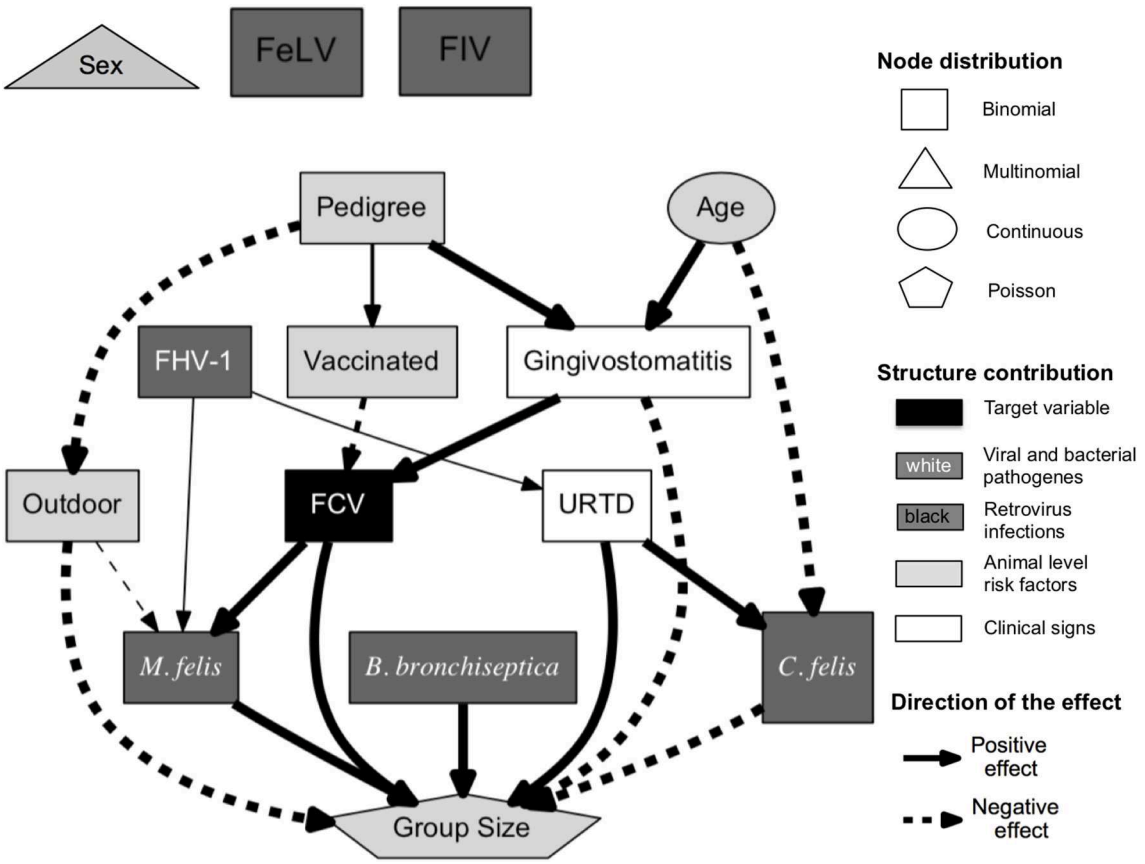


FIGURE 7 | Pruned additive Bayesian network model selected using BIC score, with a maximal parent limit of seven, constructed by pooling results across 100,000 MCMC moves [thinning factor of 100, burn-in phase of 17% of the total number of steps, and non-informative global network prior (except the already mentioned network restriction)]. The squares are binomial, the pentagon is Poisson, ovals are Gaussian, and the triangle is based on multinomial distributed variables. The FCV node is directly associated with the factors vaccination status, gingivostomatitis, *M. felis*, and group size. The solid arrows represent positive effects and the dashed arrows, negative effects. The thickness of the arrow is proportional to the support probability. The gray scale values of the nodes encode their level of contribution (target variable, viral and bacterial pathogens, retrovirus infections, clinical signs, and animal level risk factors).

70.7%. The FCV status is positively associated with gingivitis and stomatitis aggregated, with an odds ratio of 8.17 in all MCMC samples. Gingivostomatitis indicates an inflammation of the caudal and buccal oral mucosa and, occasionally, other oral mucosal surfaces. The FCV status is also positively directly associated with the presence of *M. felis*, with an odds ratio of 2.69, with a supportive probability of 100%. Housing cats in large groups is also found to be a risk factor (with a rate ratio of 1.57 present in all MCMC samples).

The original study used a dichotomized variable for group size. In the present case study, we used a Poisson-distributed variable. Based on **Figure 5**, a zero-inflated or negative binomial may have been a better choice. Unfortunately, these distributions are not available in the abn R package. Interestingly, as was found in the original publication (44), classical signs of URTD (such as nasal discharge, ocular discharge, conjunctivitis, and sneezing) are not found to be directly associated with FCV status. The reduction of the group size and vaccination are the two actionable factors found to be directly associated with the FCV status and are recommended as a measure to control FCV infection. Alternatively, the presence of gingivostomatitis or *M. felis* infection is a strong indicator of an FCV-positive status in a cat. Another nice feature of a BN analysis is the possibility of gaining insights into the relationships with variables other than the targeted one. Beyond the FCV interpretation, one can see that cats with a pedigree are more likely to be vaccinated, less likely to have outdoor access, and less likely to suffer from gingivostomatitis complex compared to non-pedigree cats. The older a cat is, the more likely it is to suffer from gingivostomatitis complex but the less likely it is to be *C. felis*-positive. **Figure 7** also shows that a cat's sex with reproductive status, FeLV status, and FIV status are not associated with the rest of the network. However, only five positive cases occur in the dataset for FeLV and FIV status, and so this result should be treated cautiously. Sex and reproductive status seem not to play a role in FCV infection dynamics.

TABLE 2 | Regression coefficient estimates and 95% Confidence Intervals (CI) with their interpretation and data support (computed with structural MCMC).

Arc	Coefficient	95% CI	Interpretation	Support [%]
FCV–Vaccinated	0.38	[0.2;0.72]	Odds ratio	70.7
FCV–Gingivostomatitis	8.17	[4.63;14.42]	Odds ratio	100
C. felis–URTD	22.20	[3.13;157.59]	Odds ratio	100
C. felis–Age	0.33	[0.14;0.77]	Odds ratio	94.9
M. felis–FCV	2.69	[1.62;4.48]	Odds ratio	100
M. felis–FHV-1	3.00	[1.54;5.58]	Odds ratio	53.2
M. felis–Outdoor	0.50	[0.31;0.82]	Odds ratio	59.9
Gingivostomatitis–Pedigree	3.00	[1.74;5.20]	Odds ratio	100
Gingivostomatitis–Age	1.54	[1.19;1.98]	Odds ratio	98.7
URTD–FHV-1	2.69	[1.41;5.14]	Odds ratio	56.3
Outdoor–Pedigree	0.12	[0.07;0.22]	Odds ratio	100
Group size–FCV	1.57	[1.35;1.82]	Rate ratio	100
Group size–*C. felis*	0.61	[0.44;0.83]	Rate ratio	97.2
Group size–*M. felis*	1.26	[1.10;1.45]	Rate ratio	95.3
Group size–*B. bronchiseptica*	2.56	[2.02;3.24]	Rate ratio	100
Group size–Gingivostomatitis	0.77	[0.66;0.90]	Rate ratio	99.1
Group size–URTD	1.27	[1.11;1.46]	Rate ratio	99.2

6. DISCUSSION AND PERSPECTIVE

This paper introduced BN modeling and highlighted its strengths and weaknesses when applied to complex epidemiological data. We illustrated the key concepts and presented a detailed case study analysis using open data and open code. We hope this will help raise awareness of BN modeling and its potential within the epidemiological community. As a secondary objective, the case study focuses on running a complimentary analysis on an already published dataset about FCV infection among cats in Switzerland. The BN modeling attempts to identify potential determinants of the FCV status and to contrast results with previous results obtained with the standard multivariate approach. The two analyses show very similar results, and the ABN analysis is a convincing alternative to the original statistical approach based on uni- and multi-variate regression models. BN modeling can be seen as a new tool that might be useful for giving additional new insights potentially not captured by classical methods.

BN modeling is typically a *hypothesis-generating* approach used in veterinary epidemiology when very little is known within a research domain. For confirmatory studies, more traditional epidemiological approaches are usually preferred. In machine learning, it is often advisable to follow guidelines for good modeling practices (51, 52). According to the literature, three major points are essential components for good modeling practices:

- Definitions of model objectives and lists of the model assumptions and algorithms used are needed
- Model outputs should be assessed
- The model's outputs must be fully reported.

Good modeling practices are essential to produce robust models, as is transparently reporting possible technical or computational issues and their workarounds. Ideally, we hope to identify and report the single most robust DAG. A significant concern in BN modeling is the tendency to overfit the data and to select overly complicated models that generalize poorly. Albeit being popular and accepted, pruning DAGs using bootstrapping leads to crude choices regarding the possible connections in the model and diminishes the range of possible interpretations. Indeed, an arc is either present or absent. This approach is somewhat rudimentary, considering the massive number of *a priori* networks. Another possible focus would be to seek robust quantification of the connection between variables among the vast number of possible models. In the case study, we emphasize the practical need to account for uncertainty in the final reported DAG through Bayesian model averaging. This methodology is a very active research field that shows encouraging results in closely related domains and seems to be the future of BN modeling. Bayesian model averaging could be very useful in an applied context to avoid reducing the richness of BN modeling to only one single model. Indeed, it allows users to quantify the marginal impact of relationships (arcs in a network) of interest by marginalizing out over networks or nuisance dependencies (i.e., all other possible relationships). Structural MCMC seems to be a very elegant and natural way to quantify the true marginal impact so that one can determine if its magnitude is great enough to consider it as a worthwhile intervention. The main drawback of this technique is its considerable computational demands. The increasing availability of cheap computational resources makes structural Bayesian model averaging feasible for a large variety of studies.

DATA AVAILABILITY STATEMENT

The datasets analyzed for this study are publicly available and can be found in the `abn` R package through the command `data("FCV", package = "abn")`.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because the veterinarians obtained informed consent from the cat owners. All samples were taken as part of a diagnostic workup or for routine testing of FeLV/FIV in healthy cats and all results were provided to the veterinarians and the cat owners; no ethical approval was necessary for this study in compliance with Swiss regulations. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

GK conceived and wrote the manuscript with support from SH, RF, and FL. GK performed the analysis of the case study. GK was co-author of the `abn` R package and creator and author of the `mcmcabn` R package. FL was the original creator and author of the ABN method, provided the critical feedback on

the structure of the manuscript and on the statistical approach used. RH-L provided the dataset, ensured the possibility of open data, and provided the critical feedback on the manuscript. BW, MM, FB, and RH-L helped with the interpretation of the case study and were the authors of the original publication. BW, FB, and PT provided the critical feedback on the manuscript. RF was the Ph.D. supervisor of GK. RF probed the abn R package and provided useful suggestions that led to numerous improvements on the package. RF provided the input on the statistical framework, model implementation, and findings reporting. SH identified the case study dataset and provided input on the findings interpretation. All authors revised the manuscript.

FUNDING

The original study (44) was financially supported by Boehringer Ingelheim and Biokema SA. The funders agreed on these additional analyses; they had no influence on the study design, execution, data analysis, or reporting of the findings.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.00073/full#supplementary-material

REFERENCES

1. Lewis F, Brülisauer F, Gunn G. Structure discovery in Bayesian networks: an analytical tool for analysing complex animal health data. *Prev Vet Med.* (2011) 100:109–15. doi: 10.1016/j.prevetmed.2011.02.003
2. Comin A, Jeremiasson A, Kratzer G, Keeling L. Revealing the Structure of the associations between housing system, facilities, management and welfare of commercial laying hens using additive Bayesian networks. *Prev Vet Med.* (2019) 164:23–32. doi: 10.1016/j.prevetmed.2019.01.004
3. Kleinbaum DG, Kupper LL, Muller KE, Nizam A. *Applied Regression Analysis and Other Multivariable Methods*. Vol. 601. Belmont, CA: Duxbury Press (1988).
4. Kratzer G, Furrer R. Varrank: an R package for variable ranking based on mutual information with applications to observed systemic datasets. *arXiv* (2018) 180407134. arXiv:1804.07134
5. Firestone SM, Lewis FI, Schemann K, Ward MP, Toribio JAL, Taylor MR, et al. Applying Bayesian network modelling to understand the links between on-farm biosecurity practice during the 2007 equine influenza outbreak and horse managers' perceptions of a subsequent outbreak. *Prev Vet Med.* (2014) 116:243–51. doi: 10.1016/j.prevetmed.2013.11.015
6. Scutari M, Denis JB. *Bayesian Networks: With Examples in R*. New York, NY: Chapman and Hall/CRC (2014).
7. Pearl J, Mackenzie D. *The Book of Why: The New Science of Cause and Effect*. New York, NY: Basic Books (2018).
8. Jewell NP. *Statistics for Epidemiology*. New York, NY: Chapman and Hall/CRC (2003).
9. Lebwohl M, Quijije J, Gilliard J, Rollin T, Watts O. Topical calcitriol is degraded by ultraviolet light. *J Invest Dermatol.* (2003) 121:594–5. doi: 10.1046/j.1523-1747.2003.12421.x
10. Pearl J. Fusion, propagation, and structuring in belief networks. *Artif. Intell.* (1986) 29:241–88.
11. Alvarez-Galvez J. Discovering complex interrelationships between socio-economic status and health in Europe: a case study applying Bayesian networks. *Soc Sci Res.* (2016) 56:133–43. doi: 10.1016/j.ssresearch.2015.12.011
12. Whitney CW, Lanzanova D, Muchiri C, Shepherd KD, Rosenstock TS, Krawinkel M, et al. Probabilistic decision tools for determining impacts of agricultural development policy on household nutrition. *Earths Future.* (2018) 6:359–72. doi: 10.1002/2017EF000765
13. Smith RL, Gallicchio LM, Flaws JA. Understanding the complex relationships underlying hot flashes: a Bayesian network approach. *Menopause.* (2018) 25:182–90. doi: 10.1097/GME.0000000000000959
14. Fenton N, Neil M. *Risk Assessment and Decision Analysis with Bayesian Networks*. New York, NY: CRC Press (2012).
15. Pittavino M, Dreyfus A, Heuer C, Benschop J, Wilson P, Collins-Emerson J, et al. Comparison between generalized linear modelling and additive Bayesian network; identification of factors associated with the incidence of antibodies against leptospira interrogans SV pomona in meat workers in New Zealand. *Acta Trop.* (2017) 173:191–9. doi: 10.1016/j.actatropica.2017.04.034
16. Ruchti S, Kratzer G, Furrer R, Hartnack S, Würbel H, Gebhardt-Henrich SG. Progression and risk factors of pododermatitis in part-time group housed rabbit does in Switzerland. *Prev Vet Med.* (2019) 166:56–64. doi: 10.1016/j.prevetmed.2019.01.013
17. Ludwig A, Berthiaume P, Boerlin P, Gow S, Léger D, Lewis FI. Identifying associations in *Escherichia coli* antimicrobial resistance patterns using additive Bayesian networks. *Prev Vet Med.* (2013) 110:64–75. doi: 10.1016/j.prevetmed.2013.02.005
18. Hidano A, Yamamoto T, Hayama Y, Muroga N, Kobayashi S, Nishida T, et al. Unraveling antimicrobial resistance genes and phenotype patterns among *Enterococcus faecalis* isolated from retail chicken products in Japan. *PLoS ONE.* (2015) 10:e0121189. doi: 10.1371/journal.pone.0121189
19. Hartnack S, Odoch T, Kratzer G, Furrer R, Wasteson Y, L'Abée-Lund TM, et al. Additive Bayesian networks for antimicrobial resistance and potential risk factors in non-typhoidal Salmonella isolates from layer hens in Uganda. *BMC Vet Res.* (2019) 15:212. doi: 10.1186/s12917-019-1965-y
20. Koller D, Friedman N. *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA: MIT Press (2009).
21. Kaplan D. *Structural Equation Modeling: Foundations and Extensions*. Vol. 10. Newbury Park, CA: Sage Publications (2008).
22. Gupta S, Kim HW. Linking structural equation modeling to Bayesian networks: decision support for customer retention in virtual communities. *Eur J Oper Res.* (2008) 190:818–33. doi: 10.1016/j.ejor.2007.05.054
23. Chickering DM. Learning Bayesian networks is NP-complete. In: Fisher D, Lenz H-J, editors. Learning from Data: Artificial Intelligence and Statistics V *Learning From Data*. Springer-Verlag (1996). p. 121–30.
24. Nagarajan R, Scutari M, Lèbre S. *Bayesian Networks in R*. New York, NY Springer-Verlag (2013). p. 125–7.
25. Verma T, Pearl J. *Equivalence and Synthesis of Causal Models*. Los Angeles, CA: UCLA, Computer Science Department (1991).
26. Friedman N, Goldszmidt M. Learning Bayesian networks with local structure. In: Jordan MI editor. *Learning in Graphical Models*. Dordrecht: Springer-Verlag (1998). p. 421–59.
27. Pearl J. *Causality: Models, Reasoning and Inference*. Vol. 29. Springer-Verlag (2000).
28. Lewis FI, McCormick BJ. Revealing the complexity of health determinants in resource-poor settings. *Am J Epidemiol.* (2012) 176:1051–9. doi: 10.1093/aje/kws183
29. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ.* (2006) 332:1080. doi: 10.1136/bmj.332.7549.1080
30. Kratzer G, Furrer R, Pittavino M. Comparison between suitable priors for additive Bayesian networks. In: Argiento R, Durante D, Wade S, editors, *Bayesian Statistics and New Generations. BAYSM 2018. Springer Proceedings in Mathematics & Statistics, Vol. 296*. Cham: Springer (2019).
31. Little RJ, Rubin DB. *Statistical Analysis With Missing Data*, Vol. 793. John Wiley & Sons (2019).
32. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the em algorithm. *J R Stat Soc B Methodol.* (1977) 39:1–22.
33. Friedman N. Learning belief networks in the presence of missing values and hidden variables. In: *ICML*, Vol. 97, Berkeley, CA (1997). p. 125–33.

34. Bernardo J, Bayarri M, Berger J, Dawid A, Heckerman D, Smith A, et al. The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures. *Bayesian Stat.* (2003) 7:453–64.
35. Madsen AL, Lang M, Kjærulff UB, Jensen F. The Hugin tool for learning Bayesian networks. In: *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*. Berlin; Heidelberg: Springer-Verlag (2003). p. 594–605.
36. Conrady S, Jouffe L. *Introduction to Bayesian Networks & BayesiaLab*. Bayesia SAS (2013).
37. Murphy K. The Bayes net toolbox for MATLAB. *Comput Sci Stat.* (2001) 33:1024–34.
38. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team (2019). Available online at: http://www.R-project.org/. (accessed January 28, 2020).
39. Scutari M. Learning Bayesian networks with the bnlearn R package. *J Stat Softw.* (2010) 35:1–22. doi: 10.18637/jss.v035.i03
40. Kalisch M, Mächler M, Colombo D, Maathuis MH, Bühlmann P. Causal inference using graphical models with the R package pcalg. *J Stat Softw.* (2012) 47:1–26. doi: 10.18637/jss.v047.i11
41. Balov N, Salzman P. *How to Use the Catnet Package*. R package version 1.15.5. (2019). Available online at: https://CRAN.R-project.org/package=catnet (accessed January 28, 2020).
42. Boettcher S, Dethlefsen C. Deal: a package for learning Bayesian networks. *J Stat Softw.* (2003) 8:1–40. doi: 10.18637/jss.v008.i20
43. Højsgaard S. Graphical independence networks with the gRain package for R. *J Stat Softw.* (2012) 46:1–26. doi: 10.18637/jss.v046.i10
44. Berger A, Willi B, Meli ML, Boretti FS, Hartnack S, Dreyfus A, et al. Feline calicivirus and other respiratory pathogens in cats with feline calicivirus-related symptoms and in clinically healthy cats in Switzerland. *BMC Vet Res.* (2015) 11:282. doi: 10.1186/s12917-015-0595-2
45. Buuren SV, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw.* (2010) 45:1–68. doi: 10.18637/jss.v045.i03
46. van Smeden M, de Groot JA, Moons KG, Collins GS, Altman DG, Eijkemans MJ, et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Med Res Methodol.* (2016) 16:163. doi: 10.1186/s12874-016-0267-3
47. Kratzer G, Furrer R. *mcmcabn: A Structural MCMC Sampler for DAGs Learned From Observed Systemic Datasets*. R Package Version 0.3 (2019). Available online at: https://CRAN.R-project.org/package=mcmcabn (accessed January 28, 2020).
48. Moffa G, Catone G, Kuipers J, Kuipers E, Freeman D, Marwaha S, et al. Using directed acyclic graphs in epidemiological research in psychosis: an analysis of the role of bullying in psychosis. *Schizophr Bull.* (2017) 43:1273–9. doi: 10.1093/schbul/sbx013
49. Kuipers J, Thurnherr T, Moffa G, Suter P, Behr J, Goosen R, et al. Mutational interactions define novel cancer subgroups. *Nat Commun.* (2018) 9:4353. doi: 10.1038/s41467-018-06867-x
50. Kuipers J, Moffa G, Kuipers E, Freeman D, Bebbington P. Links between psychotic and neurotic symptoms in the general population: an analysis of longitudinal British National survey data using directed acyclic graphs. *Psychol Med.* (2019) 49:388–95. doi: 10.1017/S0033291718000879
51. Jakeman AJ, Letcher RA, Norton JP. Ten iterative steps in development and evaluation of environmental models. *Environ Modell Softw.* (2006) 21:602–14. doi: 10.1016/j.envsoft.2006.01.004
52. Chen SH, Pollino CA. Good practice in Bayesian network modelling. *Environ Modell Softw.* (2012) 37:134–45. doi: 10.1016/j.envsoft.2012.03.012

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

METHODS
published: 21 April 2020
doi: 10.3389/fvets.2020.00176

Animal Disease Surveillance in the 21st Century: Applications and Robustness of Phylodynamic Methods in Recent U.S. Human-Like H3 Swine Influenza Outbreaks

Moh A. Alkhamis[1,2]*, *Chong Li*[2] *and Montserrat Torremorell*[2]

[1] *Department of Epidemiology and Biostatistics, Faculty of Public Health, Health Sciences Center, Kuwait University, Kuwait City, Kuwait,* [2] *Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, St. Paul, MN, United States*

Emerging and endemic animal viral diseases continue to impose substantial impacts on animal and human health. Most current and past molecular surveillance studies of animal diseases investigated spatio-temporal and evolutionary dynamics of the viruses in a disjointed analytical framework, ignoring many uncertainties and made joint conclusions from both analytical approaches. Phylodynamic methods offer a uniquely integrated platform capable of inferring complex epidemiological and evolutionary processes from the phylogeny of viruses in populations using a single Bayesian statistical framework. In this study, we reviewed and outlined basic concepts and aspects of phylodynamic methods and attempted to summarize essential components of the methodology in one analytical pipeline to facilitate the proper use of the methods by animal health researchers. Also, we challenged the robustness of the posterior evolutionary parameters, inferred by the commonly used phylodynamic models, using hemagglutinin (HA) and polymerase basic 2 (PB2) segments of the currently circulating human-like H3 swine influenza (SI) viruses isolated in the United States and multiple priors. Subsequently, we compared similarities and differences between the posterior parameters inferred from sequence data using multiple phylodynamic models. Our suggested phylodynamic approach attempts to reduce the impact of its inherent limitations to offer less biased and biologically plausible inferences about the pathogen evolutionary characteristics to properly guide intervention activities. We also pinpointed requirements and challenges for integrating phylodynamic methods in routine animal disease surveillance activities.

Keywords: human-like H3, swine influenza, evolutionary epidemiology, phylodynamics, phylogeography, disease surveillance

OPEN ACCESS

Edited by:
Salome Dürr,
University of Bern, Switzerland

Reviewed by:
Samantha J. Lycett,
University of Edinburgh, United Kingdom
Guido Alberto König,
National Agricultural Technology Institute, Argentina

***Correspondence:**
Moh A. Alkhamis
m.alkhamis@hsc.edu.kw

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *01 December 2019*
Accepted: *16 March 2020*
Published: *21 April 2020*

Citation:
Alkhamis MA, Li C and Torremorell M (2020) Animal Disease Surveillance in the 21st Century: Applications and Robustness of Phylodynamic Methods in Recent U.S. Human-Like H3 Swine Influenza Outbreaks. Front. Vet. Sci. 7:176. doi: 10.3389/fvets.2020.00176

INTRODUCTION

In the past few decades, genetic analysis of rapidly evolving pathogens has become an integral part of animal disease surveillance systems worldwide (1–4). Most current and past molecular surveillance studies of animal disease pathogens of both public health and economical importance such as influenza (5–7), foot-and-mouth disease (FMD) (8–10), and porcine reproductive

and respiratory syndrome (PRRS) (11–13) viruses are dependent on classical epidemiological and phylogenetic methods. These studies or surveillance systems used classical phylogenetic methods, including parsimony, neighbor-joining, or maximum likelihood (ML) approaches to either genotype novel emerging strains, classify viral lineages, or assess tree topologies to distinguish between novel and emerging strains (6, 7, 13). In addition, classical phylogenetic approaches were used to assess correlations between the similarities of nucleotide sequences and related epidemiological characteristics, while ignoring uncertainties associated with estimates of phylogenetic relationships, host, temporal, and spatial factors (7, 10, 11, 14). Furthermore, they investigated spatio-temporal and evolutionary dynamics of the virus isolates in a disjointed analytical framework and made joint conclusions from both analytical approaches (7, 10, 11, 14). Therefore, many of the past and current molecular surveillance studies of animal diseases have ignored that epidemiological and evolutionary dynamics of rapidly evolving viruses occur on approximately the same time-scale (15). Thus, studying them in a unified analytical framework will refine their interpretations and limit biased conclusions to subsequently improving the related molecular surveillance activities. Classical phylogenetic approaches are not capable of accounting for the uncertainties in evolutionary processes of rapidly evolving pathogens or integrating related epidemiological features into their phylogeny, which is an important advantage of Bayesian phylodynamic methods.

The Bayesian phylodynamic methods were borrowed from the field of evolutionary biology and have become a powerful tool for exploring the evolutionary epidemiology of infectious pathogens (14–17). During the last two decades, the rapid growth of pathogens' genetic data and computational resources increased the applications of phylodynamic methods in animal and human disease surveillance (17). These methods are capable of accounting for uncertainties, and uniquely integrate complex epidemiological and evolutionary processes in populations using a single Bayesian statistical framework (18, 19). This framework handles the parameters of the phylodynamic model as random variables, in which each parameter is set by a specified prior probability distribution (and a parallel inferred posterior probability distribution). Therefore, this innovative quantitative integration improved disease investigation by answering novel epidemiological questions about the evolutionary history, spatiotemporal origins, within and between-host transmission, and environmental risk factors for rapidly evolving pathogens (17). In fact, during the last decade, phylodynamic models have become well-established tools for studying the evolution of animal viral diseases specially influenza (20), FMD (17), and PRRS (21). Besides, several studies advocated for the integration of phylodynamic methods in the routinely molecular surveillance pipelines of animal diseases with the objectives of reclassifying viral genotypes, distinguishing between emerging and endemic viral strains, and selecting proper vaccine strains (17, 21–23). These approaches will provide a robust platform for guiding the allocation of resources within a surveillance system, for example, targeting emerging strains with higher evolutionary rates or hosts at high risk of generating new strains, which subsequently will reduce the economic costs of sampling, control, and prevention activities. Phylodynamic methods are implemented in many open-source statistical software packages, while the most popular user-friendly software package is formally known as Bayesian evolutionary analysis by sampling tree (BEAST) (24).

While past studies illustrated the great potential of phylodynamic tools, the methods are sensitive to the density and coverage of sequence sampling, selection of genetic regions, quality and quantity of the associated surveillance data, and prior selection for the evolutionary parameters (15, 25, 26). These limitations may result in biased posterior inferences, which subsequently lead to inaccurate or biological implausible conclusions about the evolutionary epidemiology of the pathogen under study (e.g., false divergence time or geographical origins). That said, most phylogenetic studies suffer from these inherent limitations. However, setting a thorough phylodynamic analytical pipeline, while acknowledging these limitations, can reduce their impact on the resulting posterior inferences and their related conclusions. Unfortunately, many published phylodynamic studies ignored such limitations, particularly in their analytical approach, in which they used simple naïve priors for their evolutionary parameters while ignoring the underlying assumptions for these priors (27–31). For example, prior selection should adhere to the assumption that different pathogens have unique evolutionary characteristics (14), and therefore, using the same simple prior on different pathogens will likely lead to the conclusion that such pathogens behaved similarly during their evolutionary history. Also, these studies ignored the impact of selecting different prior models on their posterior evolutionary inferences of the pathogen under study (26, 32). For example, the use of different prior models often leads to different conclusions about the geographical origins of the pathogen under study, and hence, Bayesian model selection is a critical step in phylodynamic analysis pipelines (25, 33).

There are many studies in the published literature comparing the results of phylodynamic models inferred from different gene segments or evolutionary parameters' priors (34–36). However, few studies raised concerns about the sensitivity of the results to the choice of different evolutionary models (20, 26) as well as suggested a focused phylodynamic analytical pipeline for animal disease molecular surveillance (37). Here, we demonstrate the basic principles for building a phylodynamic analytical pipeline, illustrate examples on the impact of gene segment and prior selection on the posterior evolutionary inferences, and highlight the prospects of the methods in improving animal disease surveillance. We selected a publicly available dataset compromising of 352 full genome sequences for human-like H3 swine SI collected as part of the United States Department of Agriculture influenza surveillance system between 2015 and 2018 as a working example. We provided a detailed description of a classical phylodynamic analytical pipeline encompassing both demographic and discrete phylogeographic reconstruction of the human-like H3 virus using BEAST. Our phylodynamic analyses included comparisons between commonly inferred evolutionary posterior parameters (e.g., substitution rate/site/year, divergence times, phylogeographic root state posterior probabilities, significant dispersal route between states) under different combinations of node–age and branch rate prior models. Furthermore, we extended this

analytical pipeline into comparing posterior parameters inferred from HA and PB2 gene segments. Interpretation of the resulting posterior inferences under different scenarios, described above, has been discussed in detail, and we highlighted examples of their misuse in past phylodynamic studies. Our results identified the prospects and limitations of the presented phylodynamic pipeline in the context of animal disease surveillance on regional and global scales. Furthermore, our results provide researchers and stakeholders of the swine industry in the United States valuable insights on decisions related to the sampling and sequencing of the influenza virus genome when conducting future phylodynamic studies and improving the design of currently implemented surveillance systems.

BAYESIAN PHYLODYNAMIC STATISTICAL FRAMEWORK

The summary flow chart of our phylodynamic analytical pipeline is presented in **Figure 1**. This Bayesian statistical framework is popular and well-established for studying rapidly evolving pathogens as described elsewhere (37–39). The pipeline is divided into five steps (**Figure 1**), in which two steps are dedicated to sequence preparation and curation of relevant viral lineages, while the following three steps are dedicated for phylodynamic analyses of the subsequently selected lineages.

Sequence Preparation

Sequence Collection and Retrieval

A critical step for a sound phylodynamic analysis is sequence preparation. This step can take two directions, depending on the study design and the objectives of the analysis. The first direction involves primary data analyses of novel sequences, in which they are either part of a designed study to identify the evolutionary characteristics of newly emerging viral strains (27, 37, 39) or part of an ongoing active surveillance program (40). This direction usually includes the collection and sequencing of novel viral isolates from ongoing outbreaks. The second direction involves secondary data analyses of sequence collections published in publicly available genomic databases such as the Genbank, to mainly explore the evolutionary history of specific pathogens either on regional or global scales (38, 41, 42). Secondary

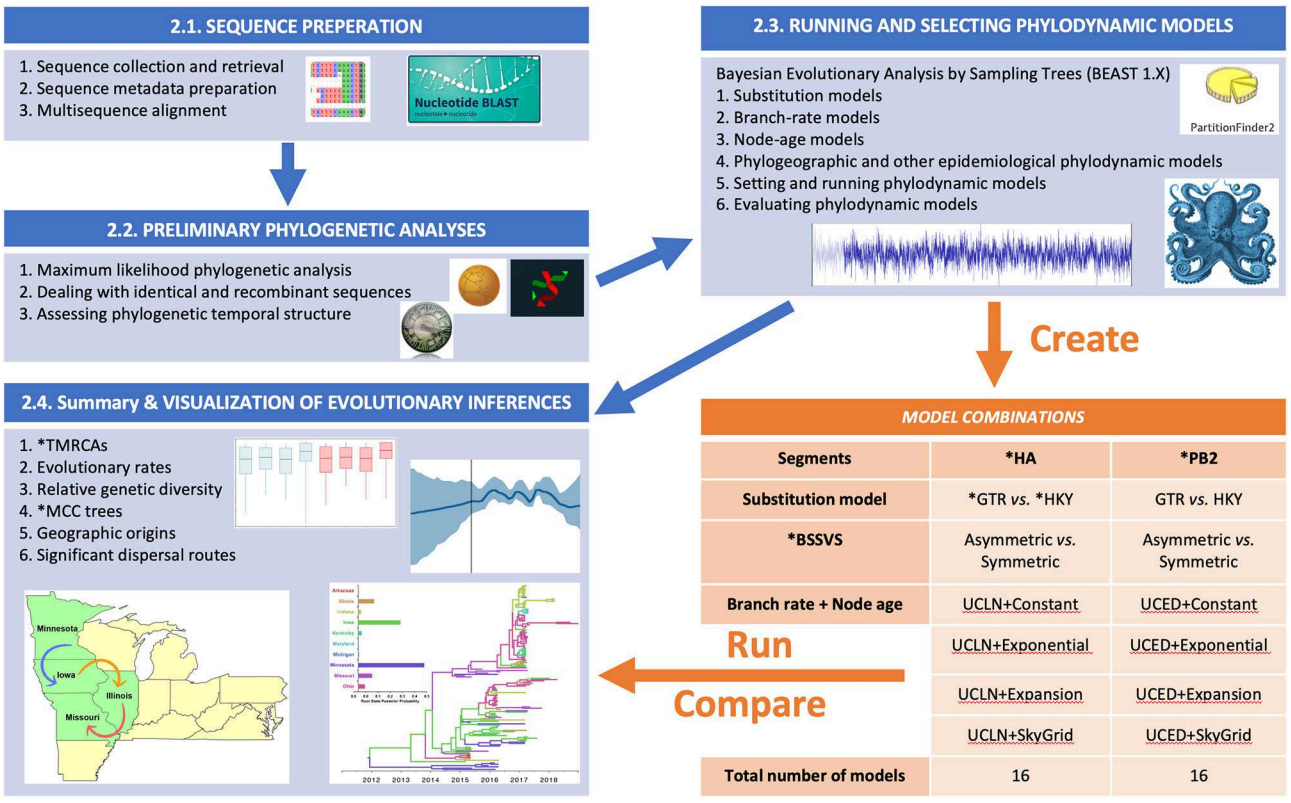


FIGURE 1 | Flow chart showing the steps for our Bayesian phylodynamic statistical framework. Blue boxes summarize the methodological steps for the statistical framework described in the BAYESIAN PHYLODYNAMIC STATISTICAL FRAMEWORK section of the main text (Sequence Preparation section to Summary and Visualization of Evolutionary Inferences section). The orange table indicates the generated models for swine influenza example data described in Section Worked Example: Evolutionary Dynamics of Swine Influenza in the United States Between 2015 and 2018. ML, maximum likelihood; RDP, Recombination Detection Program; ESS, effect sample size; BF, Bayes factor; PS, path sampling; SS, stepping stone; HA, hemagglutinin; PB2, polymerase basic 2; GTR, general time-reversable; HKY, Hasegawa, Kishino, and Yano; BSSVS, Bayesian Stochastic Search Variable Selection; TMRCAs, Time to the Most Recent Common Ancestor; MCC, Maximum Clade Credibility.

sequence analysis can either target all available viral isolates or specific well-defined lineages (i.e., monophyletic clades) (38, 41, 42). To reduce the impact of sampling bias on the results of a phylodynamic analyses, it is essential to ensure the representativeness of the viral isolates under study to the available sequences data on both temporal and spatial scales. This step is most important for primary sequence analyses, in which the dataset under study needs to cover all close relatives of novel viral isolates published elsewhere. Retrieving and combining relatives of novel viral isolates in a single dataset will warrant a proper inference of representative phylogenetic relationships of a tree topology based on all available related sequences. As on many occasions, novel sequences might belong to different distinct viral lineages published elsewhere (39, 43). The basic local alignment search tool (BLAST; https://blast.ncbi.nlm.nih.gov/Blast.cgi) is the most popular tool for retrieving relatives of novel sequences (**Figure 1**). Finally, the retrieval process should include complete and near-complete sequences to avoid distorting the phylogenetic relationships between the novel and the related isolates.

Sequence Metadata Preparation

The integration of a pathogen's epidemiological characteristics into its inferred phylogeny is the ultimate justification for the preference of the phylodynamic approach over the classical phylogenetic methods. Therefore, the thorough preparation of sequence metadata, which includes retrieval of information related to the isolate under study, is another critical step for a sound subsequent phylodynamic analysis. Sequence metadata can be retrieved either from public genomic databases such as the Genbank or from the related published literature. Because phylodynamic methods largely depend on time-stamped data, this step starts with retrieving the data of collection for the viral isolates under study. Thus, viral isolates with no temporal data are typically excluded from the analyses pipeline. Next, the date of collection is converted into BEAST readable format known as fractional years to estimate divergence times. For example, a virus collected on April 14, 2017, is converted into "2017.282" as a fractional year, where "2017" is the year of collection, and "0.282" is the number of days from the beginning of that year till the day for sequence collection divided by the total number of days within a typical year. Additionally, dates can be imported to BEAUTi by a separate text file that include the complete date of sequence collection with explicit separators (e.g., – or /). However, in many instances, the complete date of collection is not available, in which it misses either the exact date or month of collection. Therefore, we can either specify the age of the isolate as the mid-point of the corresponding month or year, respectively. Other epidemiological characteristics such spatial or host information can be prepared in a separate text delaminated format with unique identifiers that link them to the isolates in the sequence dataset. Isolates missing a none-temporal information should be kept in the analyses and are usually labeled with a question mark "?" to represent a missing information. In the context of the phylodynamic field, epidemiological characteristics such as country or host of origin are defined as a discrete trait and are described in more detail in the Running and Selecting Phylodynamic Models section. However, careful selection of these characteristics is recommended to be considered at the beginning of the analyses pipeline as a critical part of the data preparation for the subsequent analyses. Geographical discrete traits can be defined as the country of origin where the pathogen was isolated or can be redefined on smaller or larger spatial scales such as administrative regions within a country (44) or continental scale (32), respectively, depending on the study's hypothesis. Besides the host of origin, other non-spatial discrete traits such as host and environmental attributes can also be defined as discrete traits (45).

Multisequence Alignment

Multisequence alignment (MSA) is another primary key step in the data preparation stage of the pathogen's genetic data (**Figure 1**). It is worth noting that alignment uncertainty, for example, in terms of the choice of alignment algorithm can affect the subsequent phylogenetic inferences, such as tree topology (46). However, the impacts of alignment uncertainties have not been reported with simple pathogens like viruses, mainly when dealing with small gene segments. Therefore, this issue might be considered when dealing with whole genomes or with more complex pathogens like bacteria and fungi, which can be resolved by multiple sequence alignment averaging using different alignment algorithms (47). Common alignment algorithms include CLUSTAL (48), T-coffee (49), and MUSCLE (50), while AliView is a user-friendly graphical interface that can deal with large sequence datasets and integrate multiple alignment algorithms (51). Performing the multisequence alignment using an algorithm, and manually deleting the gaps within the translated alignment, are the most common steps for most phylogenetic studies (51). Also, confirming the reading frame of each gene segment (excluding the 5′UTR) by examining the amino acid translation is another step within the MSA procedure. This step is commonly done, for example, for influenza virus HA and PB2 gene segments, and potentially for segments 7 and 8, to account for the frameshifted M2 and NS2 genes. However, it is worth noting that this step is only important for partitioned nucleotide models, described below.

Preliminary Phylogenetic Analyses

Inferring Preliminary Phylogenetic Trees

Phylodynamic analyses require both time and computational resources, and therefore, conducting exploratory phylogenetic analyses using classical methods is an essential step that will ensure the proper setup of the subsequent phylodynamic models' priors. Classical methods for inferring basic phylogenetic trees (i.e., non-time-stamped trees) include the maximum likelihood (ML) (52), maximum parsimony (MP) (53), and neighbor-joining (54) algorithms. Inferring the basic phylogenetic tree of a sequence dataset will help in the preliminary assessment of the tree's topology in terms of the magnitude of structure across branches, degree of topological (in)congruence between different gene segments, and selection of lineages (in large datasets) for the subsequent phylodynamic analyses. Classical phylogenetic algorithms are implemented in many open-source software packages such as MEGA (55) and RaXML (56).

Dealing With Identical and Recombinant Sequences

The rapid spread and transmission of viral diseases during epidemics provide plenty of time for the pathogen to accumulate informative mutations in their genomes (57). Therefore, 100% identical sequences within a dataset dilute such information. Also, retrieved sequence datasets suffer from inherent redundancy due to sampling bias and issues related to the sequencing procedure (58). Hence, removing 100% identical sequences from the dataset under study, will reduce the impact of such redundancies, strengthen the tree structure, and shorten the computational time. Furthermore, if the proportion of 100% identical sequences was substantially large, it will typically lead to weaker evolutionary signals and subsequently poorer phylodynamic model convergence.

Recombination is a natural biological phenomenon of rapidly evolving viruses like influenza and occurs when viral genomes co-infect the same host cell and exchange fragments of their gene segments resulting in new viral strains (59). Ignoring recombination events in a sequence dataset may advisedly bias the inferred posterior phylogenetic relationships and, therefore, must be excluded (60). Recombination events can be detected using the Recombination Detection Program (61). However, recombination events are more often detected in whole genomes than in single-gene segments. Therefore, conducting phylodynamic analyses on whole-genome sequences only will lead to the exclusion of many isolates resulting in a substantially smaller dataset and subsequently biased inferences. Nevertheless, the occurrence of recombination events at the beginning of a novel viral outbreak might be limited.

Assessing Phylogenetic Temporal Structure

Assessing the magnitude of temporal structure in the phylogeny of the sequences data collected at different points of time is the final recommended step within the preliminary phylogenetic analyses stage (62). Here, the term "temporal structure" is defined as the measurable difference in terms of nucleotide or amino acid substitution between two genetic sequences sampled at two distinct points of time (63). Therefore, if the sequence data lacks sufficient temporal structure, then proceeding to the phylodynamic analysis may lead to biased posterior estimates and misleading conclusions (62). An interactive regression-based approach is implemented in the TempEst software package (62) to assess the strength of the association between sequences' sampling dates and genetic divergence through time. R^2 values closer to 1 than 0 estimated from a time-stamped ML tree using the root-to-tip genetic distance linear regression indicate a strong temporal structure (62). Finally, TempEst can identify incongruent sequences that are defined as outlier isolates that caused substantially more or less genetic divergence from the tip to the root than one would expect given their sampling date (62). Incongruent sequences usually result from low sequencing quality, alignment errors, laboratory adopted and vaccine strains, as well as natural biological processes such as recombination.

Running and Selecting Phylodynamic Models

Once the sequence dataset and their metadata are curated (by the past two steps, described above), we provide a variety of choices for selecting and running phylodynamic models depending on the objectives of the study. Steps involving prior specification, simulations, and summarizing posterior inferences are all implemented in the BEAST software package (24).

Substitution Models

Large evolutionary distances (i.e., substitution per site) between pairs of sequences caused by multiple substitution events through time can be underestimated when using simple distance measures (e.g., Hamming distance) (64). Hence, the distance correction technique provided by the substitution models can compensate for the underestimation of such large evolutionary distances (64). Phylogenetic tree algorithms such as the ML approach incorporates substitution models that employs continuous-time Markov chain (CTMC) models (52). CTMC models are stochastic methods that take values from a discrete state evolutionary space at random times, which is analogous to a nucleotide or amino acid substitution process, allowing for glimpsing the complete state history over the entire phylogeny where statistical inferences are drawn (52, 64, 65). Out of many available substitution models, the Hasegawa, Kishino, and Yano (HKY) (66) and the general time-reversable (GTR) (52, 67) are the most common models used to infer the phylogeny of rapidly evolving pathogens. Briefly, both substitution models assume a constant rate of evolution and have two major parameters, including a rate matrix (Q) and an equilibrium vector of base frequencies. However, the HKY model rate matrix has two exchangeability parameters, including one transition rate and one transversion rate parameters (66), while the GTR model has a symmetrical substitution rate matrix where all the exchangeability parameters are free (67).

Accommodating the rate variation across sites can be achieved by combining substitution models with site models such as the discrete gamma (Γ) model (68). However, when assuming that the evolution rate is equal to zero, the invariant site (I) model is combined with the corresponding substitution model (69). Selection pressure in protein-coding genes of rapidly evolving pathogens, in terms of synonymous to non-synonymous substitutions, usually occurs at high rates (70). This evolutionary phenomenon can affect estimates of divergence time and, therefore, need to be accounted for when selecting a substitution model (71). Partitioning the gene segment into unique codon positions and assigning different substitution and site model combinations can accommodate the differences in the evolutionary dynamics within gene segments of the pathogen under study (70, 72). Different substitution, sites, and codon partitioning models are implemented in many ML software packages as well as in BEAST. However, selecting the most realistic substitution/site model and partitioning scheme for the sequence data can be statistically achieved using either Bayesian Information Criterion (BIC) (73), Akaike Information Criterion (AIC), or the corrected Akaike Information Criterion

(AICc) (74, 75). These ML-based statistical methods are well-implemented in both PartitionFinder (76) and jModelTest (77). Yet, a more robust Bayesian method for selecting a site model and an associated substitution model is implemented as an add-on package in BEAST 2.X (78).

Branch-Rate Models

Time-calibrated trees are modeled with the genetic differences between sequences through the molecular clock models, which is defined as the clock that occurs after a stochastic waiting time in the context of substitution rate (79). When assuming that the substitution rate across the branches is uniform over the entire tree, then the molecular clock model is defined as strict. However, changes in the rate of evolution of rapidly evolving pathogens usually differ between the subtrees of its inferred phylogeny, and therefore, relaxed branch-rate models account for the variation in the rate of molecular evolution from clade to clade across the branches of the tree (79). Substitution rates across branches are assumed to be either autocorrelated (80) (i.e., substitution rates are dependent) or uncorrelated (81) (i.e., substitution rates are independent). The uncorrelated branch-rate prior commonly used for rapidly evolving viruses, in which the branch rates are drawn either from exponential or log-normal parent distribution (81). Another alternative to the strict clock model is local molecular clocks, which can estimate different rates for different predefined branch groups within a tree (82). However, for large datasets, the manual task of assigning branches to different groups is impractical (81), and therefore, Bayesian random local clocks can nest a series of local clocks with each extending over a group of branches within the full phylogeny (83).

Node-Age Models

Phylogenetic trees are inferred from individually sampled sequences to estimate the statistical properties of the population where the sequences were collected (84). Kingman's n-coalescent theory (i.e., node-age model) is the first stochastic model framework aimed at estimating the size of the sequences' population (85). The theory describes the distribution of coalescent times in the phylogeny as a function of the size of the population from which the sequences were drawn (85). Hence, in the past few decades, the coalescent theory is the core of phylodynamic methods and has shown to be the most useful for inferring essential parameters that shapes the evolution and population dynamics of evolving populations including their effective size (86), rate of growth (87), structure (88), recombination, and reticulate ancestry (89). Expanding the temporal frame of sampling times is the ultimate approach for increasing the statistical power and precision of the coalescent model in estimating substitution rates and population demographics of rapidly evolving viruses (90). An essential evolutionary parameter estimated from the coalescent model is effective population size (N_e) at a specific time (t) and interpreted as the natural population that represents sample genealogies that have statistical features of an idealized population size through time $N_e(t)$ (84). However, such interpretation is only suitable for a non-recombinant single population, whereas complex populations with more frequent recombination events require the use of structured tree models (84) described in the following section.

Estimating the posterior phylogeny of a well-mixed population with changing population size can be attained using either parametric or non-parametric node-age models (84). Parametric node-age models accommodate standard continuous population functions, the simplest and most naïve, namely, the constant population growth (CP), which assumes that the population growth rate is zero (91). The other three parametric models include the logistic (LG) growth (assumes the population growth rate is decreasing over time), exponential (EX) growth (assumes the population growth rate is fixed over time), and the expansion (EGx) growth (assumes the population growth rate is increasing over time) (91). One would expect, in the event of an epidemic caused by a rapidly evolving virus like influenza and in the absence of new vaccination, the population growth rate of the virus would realistically fit either an exponential or an expansion growth rate model (44, 92).

Unlike parametric node-age models, non-parametric models can be used to visually infer the history of population size through time (i.e., genetic diversity) from the sequence data in terms of inclines and declines (93). These models treat each coalescent interval as a separate segment to represent a parameter for population size in a given time, in which the number of segments can be specified by the investigator to generate a sky plot (93). The piece-wise constant Bayesian skyline (BS) is the simplest non-parametric model, which assumes that the effective population size is experiencing an episodic stepwise changes through time (93). However, the BS model is shown to be very sensitive to the total number of change points (i.e., coalescent intervals) when specified as a prior as well as to the number of sequences sampled at each point of time (94). Hence, a Gaussian Markov random fields Bayesian Skyride (GMRF) was proposed as an alternative model to BS (95). The GMRF model is less sensitive to the prior number of change points because it implements a temporal smoothing approach to recover accurate population size trajectories (95). However, an improved version of the GMRF is the Skygrid (SG), which takes into account mutation parameters of multi-locus sequences (33). The SG provides a more realistic estimate of demographic history in terms of population size and divergence times, as well as flexibility in terms of the ability to specify cut-points to the time trajectories (33). Furthermore, the SG model is the least sensitive to the temporal distribution of sequences (33). A notable example of sky plots utility in PRRS virus molecular surveillance in the United States was demonstrated by Alkhamis et al. (21) and Alkhamis et al. (37). Their sky plot inferred a distinctly high genetic diversity through time for the emerging 1-7-4 RFLP-type PRRV virus (37), while inferred consistent seasonal increases and decreases in the relative genetic diversity through time for endemic strains isolated between 2014 and 2015 (21).

Phylogeographic and Other Epidemiological Phylodynamic Models

Mugration models are substitution models used to infer the migration processes of evolving organisms (96). The most notable implementation of a migration model was developed by

Lemey et al. (97) using a CTMP to infer H5N1 avian influenza virus's global origins and movements between countries. They used countries from which the sequences have been sampled as discrete traits to estimate migration rates between pairs of predefined sets of geographical locations, and therefore, the method is named discrete phylogeography (97). Also, the method is known as discrete trait analysis (DTA) because it has the flexibility to use any other discrete trait such as host or farm characteristics from which the sequences have been isolated to model migration rates between infected hosts and farms (37, 98). Besides, the method can infer ancestral origins (i.e., from the assigned discrete traits) for the internal nodes of the phylogeny through their estimated root state posterior probabilities (RSPP) (97). However, the most notable feature of discrete phylogeographic models is the integration of a Bayesian stochastic search variable selection (BSSVS) procedure to identify significant viral dispersal routes between geographical regions or host species (97). BSSVS can also infer the significance of the directionality in the migration process between pairs of discrete traits through integrated symmetric and asymmetric substitution models. The symmetric (Sym) model assumes that the transition rate from state "A" to "B" is the same as the transition rate from state "B" to "A" (i.e., directional spread between traits is insignificant), while the asymmetric (Asym) model assumes that the transition rate from state "A" to "B" is different from the transition rate from state "B" to "A" (i.e., directional spread between traits is significant) (97). However, the lack of a sufficient number of sequences closer to the root of the phylogeny can impact accurate estimation of ancestral traits (i.e., ancestral geographical location or host) by the DTA method (97). Therefore, DTA robustness can be improved by increasing the geographical density and temporal depth of sampling (96). DTA is also limited by the type and number of variables that can be used to estimate ancestral states. Therefore, the BSSVS framework has been extended to accommodate a transitional rate matrix between discrete traits as a generalized linear model (GLM) (22, 32). The method improves biological plausibility of the inferred RSPP for the ancestral traitsby simultaneously estimating the inclusion probabilities of geographic, demographic, and environmental predictors (22). However, the method is shown to be more sensitive to sampling bias than the standard BSSVS approach (32). Hence, comparative sensitivity analyses to sampling bias between the approaches are recommended to avoid severely biased inferred RSPPs.

In some settings, geographical boundaries cannot be defined by discrete spatial traits such as the distribution of wildlife hosts or disease vectors and, therefore, viral evolution and spread better modeled by continuous spatial diffusion models (96). When precise geographical information is available (i.e., longitude and latitude), continuous phylogeographic can reconstruct the viral spatio-temporal evolutionary history using relaxed random walk models (19). These models can additionally estimate viral dispersal rate in km^2/year and can distinguish whether the spatial diffusion process was homogenous (e.g., dispersal by air) or heterogeneous (dispersal by movements) (19, 21).

In many instances, sequence samples tend to cluster within a geographical region leading to incomplete mixing and formation of structure in the population. This might bias the posterior inferences that estimated the coalescent phylogeographic models mentioned above. Hence, the recently developed structured coalescent tree models for inferring phylogeography can simultaneously model the migration process between regions while allowing for those regions to have their unique coalescent rates (96, 99). Unlike BEAST 1.X, BEAST 2.X has recently implemented several structured coalescent models for inferring geographic and between-host transmission histories, including Bayesian structured coalescent approximation (BASTA) (26), structured coalescent transmission tree inference (SCOTTI) (100), and marginal approximation of the structured coalescent (MASCOT) (101).

The complexity of infectious disease transmission dynamics pushed the capacity of phylodynamic models beyond demographic and phylogeographic reconstructions into investigating traditional and new epidemiological problems. One notable example was demonstrated by Volz et al. by developing a structured coalescent susceptible-infected-recovered (SIR) model to infer reproductive numbers from viral sequences data (102). Similar, but more complex, implementations of mathematical epidemiology in the phylodynamic models were described elsewhere (103, 104).

Setting and Running Phylodynamic Models

Prior phylodynamic models described above can be readily selected and set using a graphical user interface (GUI) implemented within the BEAST software package, namely, the Bayesian Evolutionary Analysis Utility (BEAUti) (24, 105). After selecting and setting the models, the software generates a standard XML format structured text file allowing for flexible modifications for more sophisticated evolutionary models. However, the generated XML files are very complex in their structure, and therefore, manual modifications should be made by relevant experts to avoid the introduction of significant error into the model (105). Additional tutorials on selecting and setting evolutionary models using BEAST 1.X are available elsewhere (106–108).

Phylodynamic model selection is a critical component of the analysis pipeline described in **Figure 1**, simply because different pathogens or gene segments have different evolutionary processes. Therefore, using a single phylodynamic model with similar priors to infer the evolution of multiple pathogens may be biologically implausible, leading to biased inferences. Exploring the fit of the sequence data to different phylodynamic model combinations, in terms of substitution, branch rate, and node age to infer divergence times, Time to the Most Recent Common Ancestor (TMRCAs), evolutionary rates is the best strategy for ensuring accurate estimation of posterior inferences. For inferring viral demographic history, our suggested pipeline (**Figure 1**) leads to the generation of eight phylodynamic model combinations for a single gene segment, including the selected substitution model (by PartitionFinder), two branch rate priors (UCED and UCLN), and four node-age priors (Cp, Ex, Exg, and SG). However, when inferring phylogeographic history using DTA, we suggest exploring both Sym and Asym BSSVS models (**Figure 1**), which will lead to the generation of 16 models. Our

rigorous analytical pipeline is indeed timely and computationally demanding, but on the other hand, it will lead to the selection of the most realistic model that fits the sequence data with confidence. However, this suggested pipeline is not a strict set of procedures that will ensure appropriate inferences, and therefore, researchers may explore other model or analytical pipelines relevant to their evolutionary hypotheses. It is worth noting that the computational efficiency has been substantially improved in BEAST version 1.10 and the accompanied software library Broad-platform Evolutionary Analysis General Likelihood Evaluator (BEAGLE; permits flexible parallel computing) when compared to earlier versions. The fit of the sequence data to the most realistic phylodynamic model can be assessed through simultaneous estimating the marginal likelihood (MLL) using the path sampling (PS) (25) and stepping-stone sampling (SS) (109) implemented in BEAUti using the standard settings (i.e., simulating across 100 samples for 1 million cycle from the posterior to the prior with a prior reflection point of Beta [0.3, 1.0]). The joint posterior probability density of the models' parameters is estimated by the MCMC algorithms. Setting the appropriate length of the MCMC chains (i.e., number of cycles) to ensure model convergence is dependent on the number of sequences in the dataset. One recommended approach is to quadratically increase the chain length relative to the number of sequences (e.g., 4 million states per sequence) (110). Finally, creating duplicate runs from each generated model can aid in assessing the performance stability of the MCMC simulations and their MLL estimates.

Evaluating Phylodynamic Models

MCMC log-files generated by BEAST can be thoroughly evaluated using a friendly GUI software known as Tracer (111). The software provides a simultaneous platform for summarizing and visualizing posterior estimates. Appropriate model convergence can be evaluated by examining the MCMC mixing (based on acceptance ratios) using trace plots, after discarding the 10% of the sample (the "burn-in"). Besides, assessing the estimates of the effective sample sizes (ESS) for each parameter, in which ESS values >200, indicates good model convergence (111). On some occasions, good model convergence does not ensure consistent parameter estimation due to the use of non-informative priors implemented in BEAUti. Therefore, it is critical to compare posterior parameter estimates (e.g., evolutionary rates, population growth rates, PS, and SS MLL estimates) between independent runs for each model to warrant that each parameter is closely identical to its duplicate run. In case of improper model convergence and inconsistent parameter estimation, it is recommended to either increase the length of the MCMC chain or the use of informative priors from previous MCMC runs for the same gene segment or pathogen.

Model selection is achieved by comparing the Bayes factor (BF) of the resulting MLL estimates (from the PS and SS methods) of their corresponding candidate models (25). Briefly, the BF value of the candidate models is summarized using a matrix and computed using the following equation:

$$BF = 2(lnp\,(Y|M_i) - \ lnp\left(Y\middle|M_j\right))$$

where Y is the sequence data, M_i is the candidate model "i," M_j is the competing candidate model "j," and $\ln p\,(Y|M)$ is the MLL estimate by either SS or PS simulators. BF values estimated by the SS method are summarized on the upper off-diagonal of the matrix, while BF values estimated by the PS method are summarized on the lower off-diagonal of the matrix. A model with horizontal (i.e., row side of the maxtrix) BF values greater than other candidate models is selected. Additional applied examples on model selection using BEAST 1.X are available elsewhere (106–108). The ultimate goal of the model selection procedure is to find the best fitting model that generated the data, while combining simplicity with biological realism, to appropriately represent the evolutionary characteristics of the pathogen under study (25, 112).

Summary and Visualization of Evolutionary Inferences

Inferred relative genetic diversity through time (or other reconstructed demographic trajectories) and its highest posterior density (HPD) interval can be summarized using sky plots (e.g., Skygrid plot) generated by Tracer. Similarly, estimates of divergence time, TMRCAs, and substitution rate/site/year with their HPD intervals can be summarized in Tracer using either box or violin plots (111). Also, Tracer provides a flexible platform for simultaneous comparison of evolutionary estimates inferred by multiple phylodynamic models.

Next, the resulting marginal posterior probability density of the selected model is summarized as a maximum clade credible (MCC) tree using TreeAnotator (24) to generate a tree file. MCC tree (from the tree file) can be then visualized and annotated with either posterior support values or RSSPs of the discrete traits at the internal nodes using FigTree (113). In addition, FigTree provides many customizable tree visualization options as well as it allows the users to upload additional information using a text file to annotate flexibly descriptions on the nodes and branches of the trees.

SpreaD3 is an interactive Java-based parsing and rendering tool that can summarize and visualize phylodynamic reconstructions to infer spatio-temporal and trait evolutionary history (114). Also, SpreaD3 integrates JavaScript D3 libraries to provide a web-based visualization platform for phylogeographic trees and their related inferences by combining information from the MCC tree and GeoJSON-based geographic map files (114). SpreaD3 can generate a time-lapse that superimposes the MCC tree annotated with either discrete or continuous spatial traits on a map, which can be visualized using either GIS-KLM virtual globe software (e.g., Google Earth) or modern web-browsers (e.g., Safari or Chrome). This time-lapse demonstrates the epidemic reconstruction of pathogen evolutionary history through space and time, which can quantify the diffusion processes within and between geographical regions. Furthermore, SpreaD3 can identify and plot well-supported rates between pairs of discrete traits using BFs estimated from the symmetric or the asymmetric BSSVS models. Statistically significant rates with large BF values can be used to demonstrate critical viral dispersal routes between geographical regions or transmission cycles between host species.

WORKED EXAMPLE: EVOLUTIONARY DYNAMICS OF SWINE INFLUENZA IN THE UNITED STATES BETWEEN 2015 AND 2018

Sequence Data

The spillover of H3 SI virus from humans to swine in the early 2010s in the United States resulted in a novel emerging virulent strain, which was antigenically distinct from endemic swine strains, and therefore was named "human-like" H3 virus (115). Swine-related anthropological activities such as pig movement and vaccination are the most likely factors for the continuous emergence of SI novel strains (6). Therefore, integrating phylodynamic methods with influenza surveillance systems may reduce the continuous evolutionary implications of SI viruses on both public and animal health in the United States and worldwide. Here, we chose DTA models for our comparative phylodynamic analyses example, due to their popularity, ease of use, interpretation, and computational efficiency when compared to more complex similar models.

Hence, we retrieved HA and PB2 nucleotide sequences of human-like H3 SI from the Influenza Research Database (116) to explore their evolutionary history using our suggested phylodynamic pipeline, described above (**Figure 1**). The data comprised 352 sequences with complete date and geographical information for each gene segment and was collected from 17 U.S. states (Arkansas, Illinois, Indiana, Iowa, Kentucky, Maryland, Michigan, Minnesota, Missouri, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, South Dakota, West Virginia, Wisconsin) between January 8, 2015 and June 1, 2018. The sequence data were collected from the swine production systems and exhibition swine agricultural state fairs as part of the United States Department of Agriculture (USDA) swine influenza surveillance program (40) and was partially analyzed by Walia et al. using classical phylogenetic methods (6). We aligned the sequences for both gene segments and assessed the topological (in)congruence of their phylogeny by performing an ML analysis for the individual segments using the GTR + Γ substitution model, which entailed 10 through bootstrap searches with 100 ML replicates in each run (**Supplementary Figure 1**). For the subsequent phylodynamic analyses, we removed recombinant and 100% identical sequences, which reduced the dataset to 142 sequences for each gene segment (**Supplementary Table 1**). We then evaluated the fit of the sequences to the most realistic substitution model and partitioning scheme using the BIC approach. Finally, we evaluated the temporal signal in the sequence data and found that both segments were suitable for the subsequent molecular clock analyses ($R^2 = 0.65$ and 0.40 for HA and PB2, respectively) (**Supplementary Figure 2**).

Comparative Phylodynamic Analyses

We assessed the sensitivity of the inferred posterior evolutionary of human-like H3 SI sequence data to the choice of different gene segments (i.e., HA vs. PB2) and phylodynamic priors, including substitution, discrete spatial trait, branch rate, and node-age models on the (**Figure 1**). For each gene segment, we generated 16 phylodynamic models (a total of 32 runs for both segments) using the default none-informative priors' combinations implemented in BEAUTi (**Figure 1**). These prior models included: (1) the GTR + Γ vs. the HKY + Γ for the site models; (2) the symmetric vs. asymmetric for discrete spatial models; (3) the UCLN vs. UCED for the clock models; and (4) the CP vs. The EG vs. The EGx vs. the SG for the coalescent tree models (**Figure 1**). We excluded spatial traits (i.e., U.S. states) with only one sequence (**Supplementary Table 2**) leading to the inclusion of 10 states in the subsequent DTA. Also, we evaluated the fit of the 16 phylodynamic models to the HA and PB2 sequences using the BF comparisons of their MLL estimated by the PS and SS simulator in order to select the most realistic model and correctly interpret its posterior inferences. We then used two replicate MCMC simulations for 150 million cycles and sampled every 1,500th state for each candidate model.

After assessing for proper model convergence, we compared the inferred evolutionary demographics of each candidate model by summarizing their inferred divergence times, substitution rates, and TMRCAs. Besides, we then generated the SG plots to compare relative genetic diversity for HA and PB2 gene segments inferred from the two different sites and discrete spatial models. Similarly, we compared the phylogeographic inferences of each model by generating MCC trees, summarizing the RSPPs of the states, and plotting them at the internal nodes of their corresponding trees. Finally, we selected and plotted the statistically significant dispersal routes between states under each candidate model using a cutoff BSSVS-BF ≥ 10.

Results

Demographic Posterior Inferences of HA and PB2 Gene Segments

The BIC values, described above, indicated that the HKY + Γ is the best fitting substitution model for the HA gene segment (BIC = 13,399), while the GTR + Γ is the best fitting substitution model for the PB2 gene segment (BIC = 20,029). In addition, results of the BF values (≥5) indicates that the best fitting branch-rate and node-age models to the sequence data were the SG + UCLN for HA and SG + UCED for PB2 segments (**Supplementary Tables 3–6**). However, there were no significant changes in the posterior demographic inferences when choosing the opposite substitution model for both gene segments. Similarly, our results indicate that the choice of discrete spatial and node-age models does not substantially change the estimated divergence times and substitution rates/site/year (**Figure 2**) for each gene segment alone. Additionally, these estimates were also not sensitive to the choice of branch-rate models (i.e., UCED and UCLN). However, when comparing divergence times between segments, our results indicate substantial differences in a magnitude of ∼8 years, in which the divergence time for the HA segment was around 2013 (**Figure 2A**), while for the PB2 segment, it was around 2005 (**Figure 2C**). No differences were observed in the substitution rates/site/year between the two gene segments, which were ranging between 3.3×10^{-3} (95% HPD; from 2.8×10^{-3} to 3.9×10^{-3}) and 2.9×10^{-3} (95% HPD; from $2.2 \times$

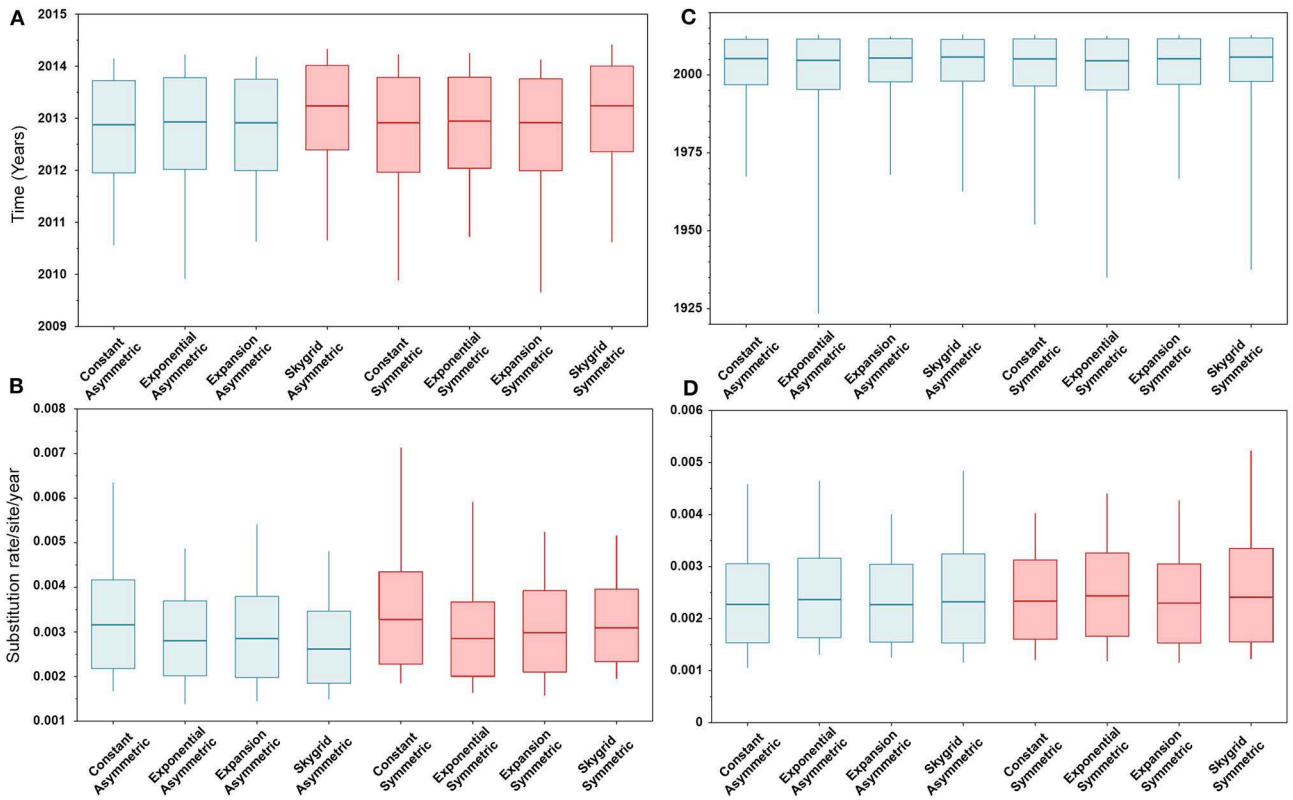


FIGURE 2 | Box plots of divergence times and evolutionary rates (substitution rate/site/year) of hemagglutinin (HA) and polymerase basic 2 (PB2) gene segments of human-like H3 swine influenza virus collected between January 2015 and June 2018 in the United States. The boxplots summarize the posterior estimates of eight phylodynamic model combinations (node-age and BSSVS priors) for each gene segment. Boxes represent 95% high posterior density (HPD), and midlines indicate the posterior median for each estimate. Blue and red boxes indicate symmetric and asymmetric models, respectively. **(A)** Divergence times of HA gene. **(B)** Substitution rates/site/year of HA gene. **(C)** Divergence times of PB2 gene. **(D)** Substitution rates/site/year of PB2 gene.

10^{-3} to 3.8×10^{-3}) for HA and PB2 segments, respectively (**Figures 2B,D**).

Similarly, posterior estimates of TMRCAs were not sensitive to the choice of phylodynamic priors but were different between the two gene segments (**Figure 3**). Hence, based on the HA segment, our results hint that the oldest human-like H3 strains emerged from the state of Minnesota in mid-2013 (**Figures 3A,B**), but with a notable overlap in the 95% HPD of the TMRCAs inferred for other states (excluding Maryland). However, results distinctly suggest that the youngest strains emerged from the state of Maryland in early 2017. Results of the PB2 segment were inconclusive in terms of determining the oldest strains, but identical to the HA gene in identifying Maryland as the state of the youngest viral strains (**Figures 3C,D**). Also, the choice of spatial trait model did not affect our estimates of genetic diversity for both HA and PB2 segments (**Figure 4**). Our SG plots inferred seasonal variations in terms of increases and decreases, in the genetic diversity through time for HA segments (**Figures 4A,B**), while the genetic diversity of the PB2 segment gene slightly declined after 2015 (**Figures 4C,D**).

Phylogeographic Posterior Inferences of HA and PB2 Gene Segments

Our inferred phylogeographic posteriors did not show sensitivity to the selection of substitution or molecular clock priors. However, substantial differences were inferred when selecting different node-age and discrete spatial trait priors. Inferences from both the CP and the EX node age with the asymmetric models implicated Missouri as the most likely ancestral state for the human-like H3 virus currently circulating in the United States when using the HA gene segment (**Figures 5A,B**). However, the EGx and the SG with the asymmetric models Illinois and Minnesota as the most likely ancestral states, respectively (**Figures 5C,D**). Yet, when using the HA segment, the symmetric model with the CP, EG, and EGx priors consistently implicated Minnesota with approximately similar estimates of RSPPs (**Figures 5E–G**). In contrast, the use of the symmetric model with the SG prior implicated Iowa as the ancestral location for the currently circulating human-like H3 strains (RSPP = 0.36) (**Figure 5H**). Interestingly, the HA sequence data uniquely favored this prior combination when using the

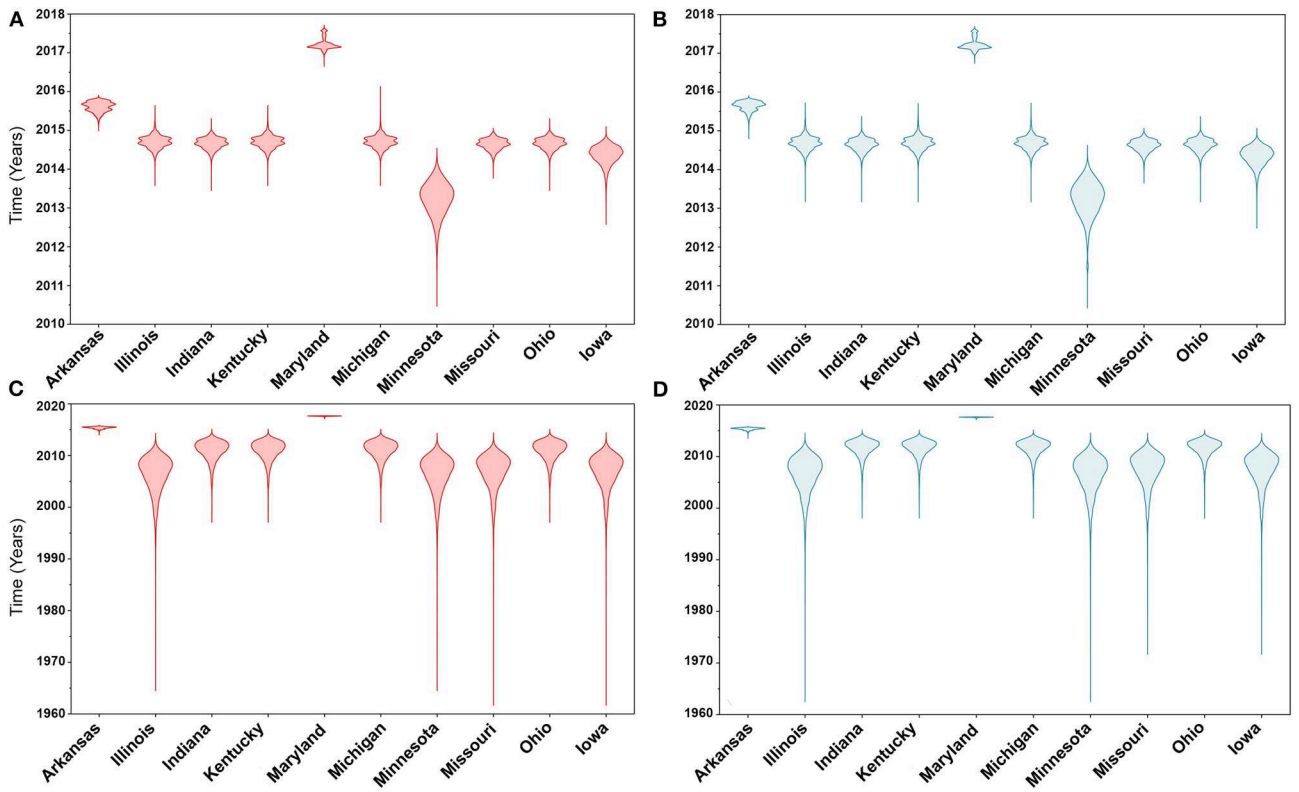


FIGURE 3 | Violin plots of state level time to the most recent common ancestors (TMRCA) of hemagglutinin (HA) and polymerase basic 2 (PB2) gene segments of human-like H3 swine influenza virus collected between January 2015 and June 2018 in the United States. The plots were generated from the Skygrid coalescent tree model. Red and blue colors indicate the asymmetric and symmetric BSSVS priors, respectively. **(A)** TMRCAs of HA gene estimated from the asymmetric BSSVS model; **(B)** TMRCAs of HA gene estimated from the symmetric BSSVS model; **(C)** TMRCAs of PB2 gene estimated from the asymmetric BSSVS model; **(D)** TMRCAs of PB2 gene estimated from the symmetric BSSVS model.

BF comparisons for the best fitting phylodynamic model (**Supplementary Tables 3**, **4**).

Our BF values suggested that the PB2 sequence data favored the asymmetric model with the SG prior, but with a very slight edge over the symmetric model with the same coalescent prior (**Supplementary Tables 5**, **6**). RSPPs inferred from the PB2 segment were almost equal for all states and, hence, were inconclusive, when using the asymmetric model with the four coalescent priors (**Figures 6A–D**). Similarly, using the symmetric model with the four coalescent priors was inconclusive in terms of identifying the ancestral location for the currently circulating viral strains (**Figures 6E–H**). More specifically, the magnitude of differences between Minnesota and Missouri and in the inferred RSPPs, across different coalescent priors, was substantially small (**Figures 6E–H**). For example, when using the SG prior, the inferred RSPPs were 0.18 and 0.22 for Missouri and Minnesota, respectively (**Figure 6H**).

Our BF-BSSVS analyses, using the asymmetric model with the CP and the EX coalescent priors for the HA gene segment, suggest that the top three most significant unidirectional routes of viral dispersal (BF $>$ 18) were between Minnesota, Iowa, Illinois, and Missouri (**Figures 7A,B**). The inferred routes maintained their unidirectionality from the origin to the destination geographical locations, using CP and EX priors (**Figures 7A,B**). Similarly, the order of statistical significance suggests that the route from Iowa to Minnesota is the most important for viral dispersal between states (**Figures 7A,B**). In contrast, the EXg with the asymmetric model suggests that the route from Ohia to Indiana is substantially the most significant dispersal route (BSSVS-BF $=$ 1,157) (**Figure 7C**). Nevertheless, the SG prior agrees with the results of the CP and EX priors in inferring the route from Iowa to Minnesota as the most significant (BSSVS-BF $=$ 37) (**Figure 7D**), while inferences from the symmetric model and the four coalescent priors consistently agreed that the top most significant bidirectional route of viral dispersal (BF $\geq$ 990) was between Indiana and Ohio (**Figures 7E–H**). However, disagreements were inferred on the second and the third most significant routes when using the CP and EX on one side and EXg and SG on the other (**Figure 7H**).

Dispersal routes inferred for PB2 (including the order of significance) were also sensitive to the selected discrete spatial model and slightly to the coalescent priors (**Figure 8**). Thus, when using the asymmetric model, the top two unidirectional routes included (1) Iowa $\rightarrow$ Minnesota; (2) Indiana $\rightarrow$ Kentucky (**Figures 8A–D**). While the CP, EX, and EXg inferred the route from Illinois to Missouri as the third most significant route

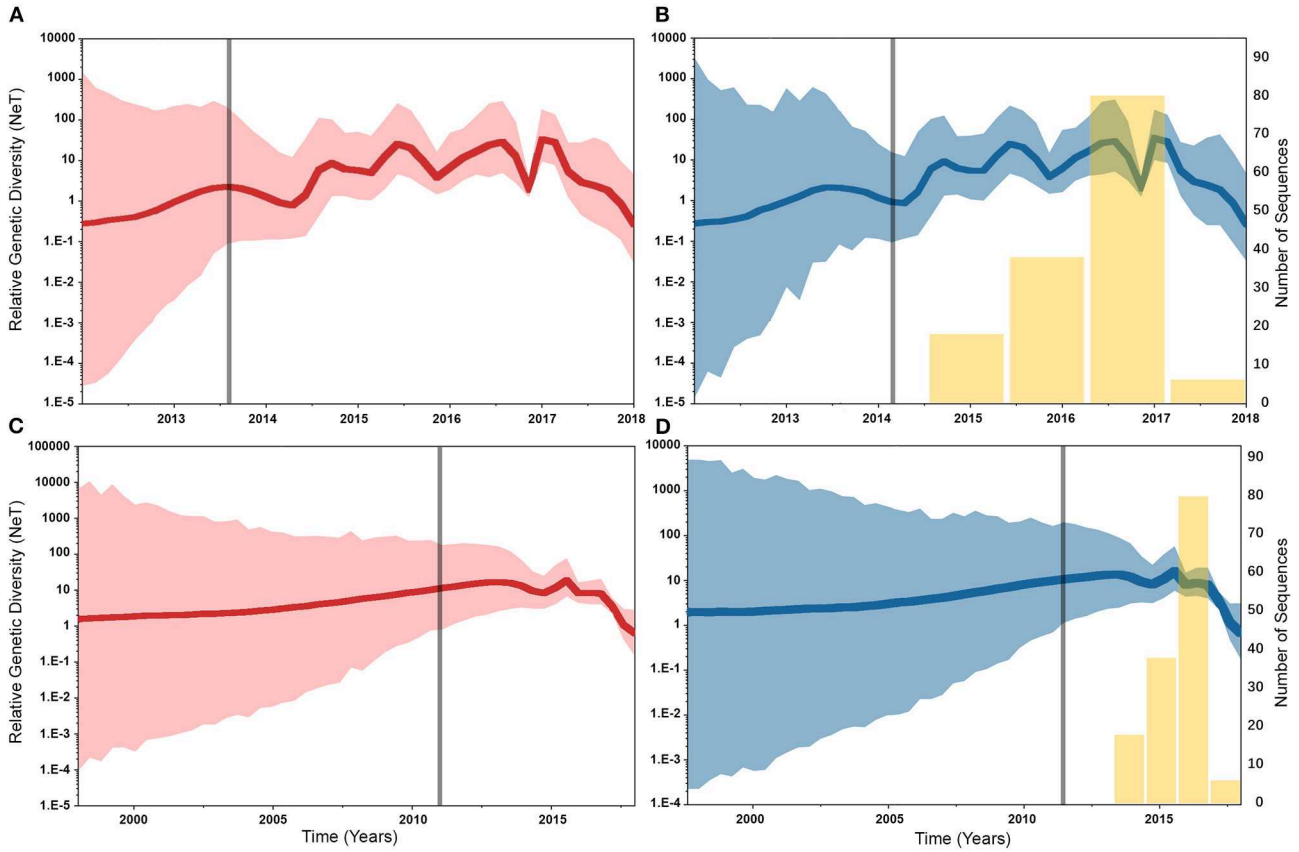


FIGURE 4 | Bayesian Skygrid (SG) plots of the relative genetic diversity through time of hemagglutinin (HA) and polymerase basic 2 (PB2) segments of human-like H3 swine influenza virus collected between January 2015 and June 2018 in the United States. The dark line indicates the posterior median estimate, and the 95% high posterior density (HPD) is indicated by the light shaded areas (red and blue for the asymmetric and symmetric BSSVS priors, respectively). The vertical gray line indicates the estimated time at which the relative genetic diversity transitioned from a slow to a fast growth rate. Yellow bars indicate the temporal distribution of the sequence data. **(A)** SG plot of HA gene inferred from the asymmetric BSSVS prior. **(B)** SG plot of HA gene inferred from the symmetric BSSVS prior. **(C)** SG plot of PB2 gene inferred from the asymmetric BSSVS prior; **(D)** SG plot of PB2 gene inferred from the symmetric BSSVS prior.

(**Figures 8A–C**), the SG prior inferred the route from Ohio to Indiana as the third most significant route (**Figure 8D**). Finally, our inferred top three significant dispersal routes were from the symmetric model between (1) Indiana and Ohio; (2) Minnesota and Iowa; (3) Indiana and Kentucky (**Figures 8E–H**).

DISCUSSION

In the past decade, our phylodynamic pipeline became well-established and demonstrated powerful potentials to trace the evolutionary history of both animal and human pathogens making it an ideal tool for designing new molecular surveillance systems. In this study, we revisited essential concepts and definitions within the field of phylodynamic methods. Also, we challenged the robustness of the posterior evolutionary parameters, inferred by the commonly used phylodynamic models, using two gene segments, of the currently circulating human-like H3 SI viruses isolated in the United States, and multiple priors. Subsequently, we compared similarities and differences between the posterior parameters inferred from HA and PB2 sequence data using multiple phylodynamic models. Hence, we explored the robust and sensitive aspects of SI phylodynamic models and highlighted the importance of model selection within their analytical framework. However, unlike classical phylogenetic methods currently implemented within the SI surveillance system in the United States, we were able to reveal higher resolution insights into the evolutionary epidemiology of human-like H3 viruses by quantifying their demographic and phylogeographic history. Therefore, animal health researchers and stakeholders need to be aware of the method's features, strengths, and limitations for generating reliable inference to guide future disease intervention activities properly.

Updated Insights in the Evolutionary Epidemiology of Swine Influenza in the U.S.

Based on the results of the best fitting phylodynamic models for both HA and PB2 segments, evolutionary rates of currently circulating human-like H3 viruses in the United States remain high with no apparent signs of substantial declines (**Figures 2B,D**) and were similar to what was inferred elsewhere

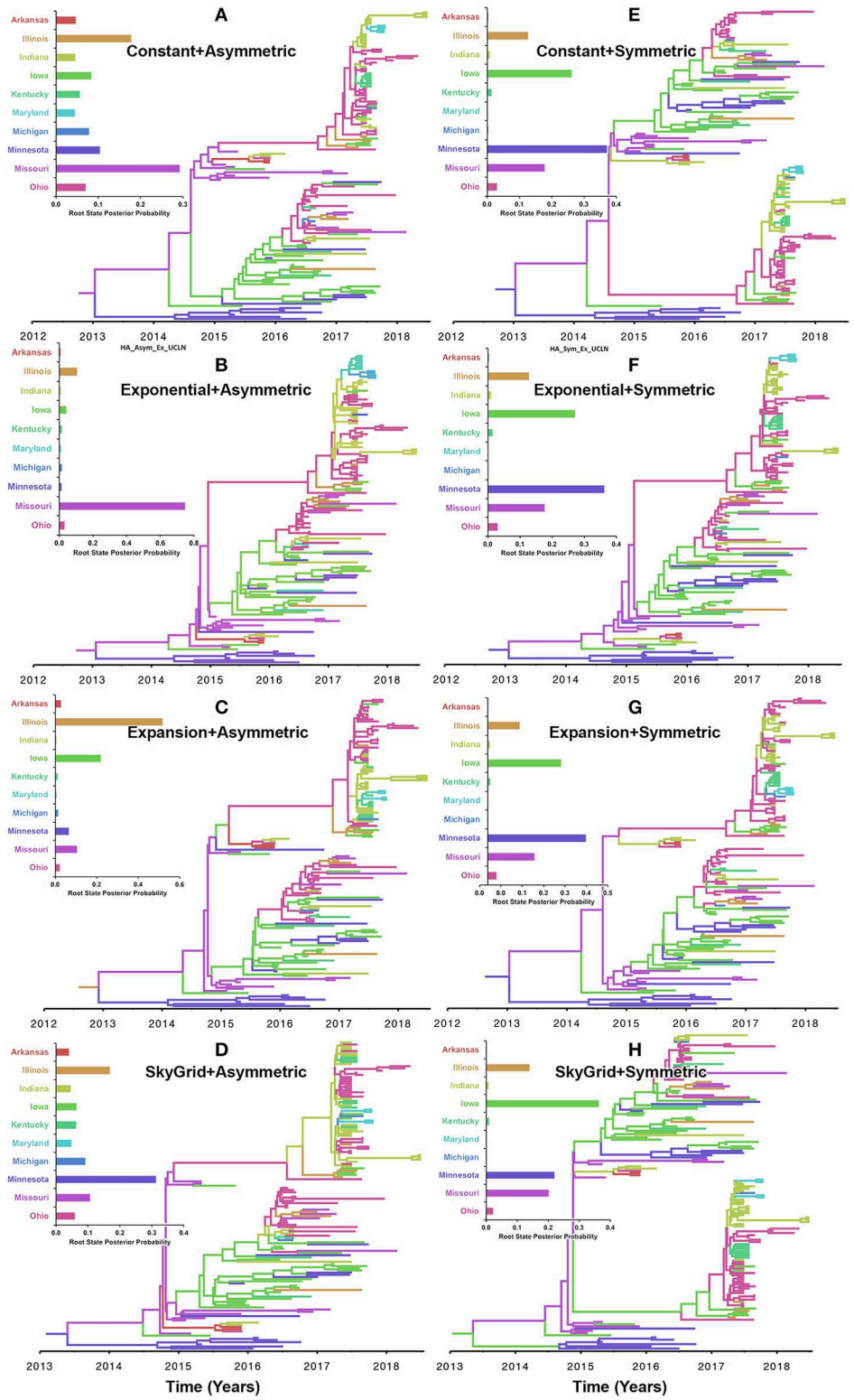


FIGURE 5 | Maximum clade credibility (MCC) phylogeny of the HA segment of human-like H3 swine influenza virus collected between January 2015 and June 2018 in the United States. The trees are inferred from eight phylodynamic model combinations (node-age and BSSVS priors). The color of the branches represents the most probable location state of their descendant nodes, and their color-coding corresponds to the upper left bar chart, which represents the root location state posterior probabilities (RSPP) for each state. **(A–D)** Trees inferred from four node-age + asymmetric BSSVS priors. **(E–H)** Trees inferred from four node-age + symmetric BSSVS priors.

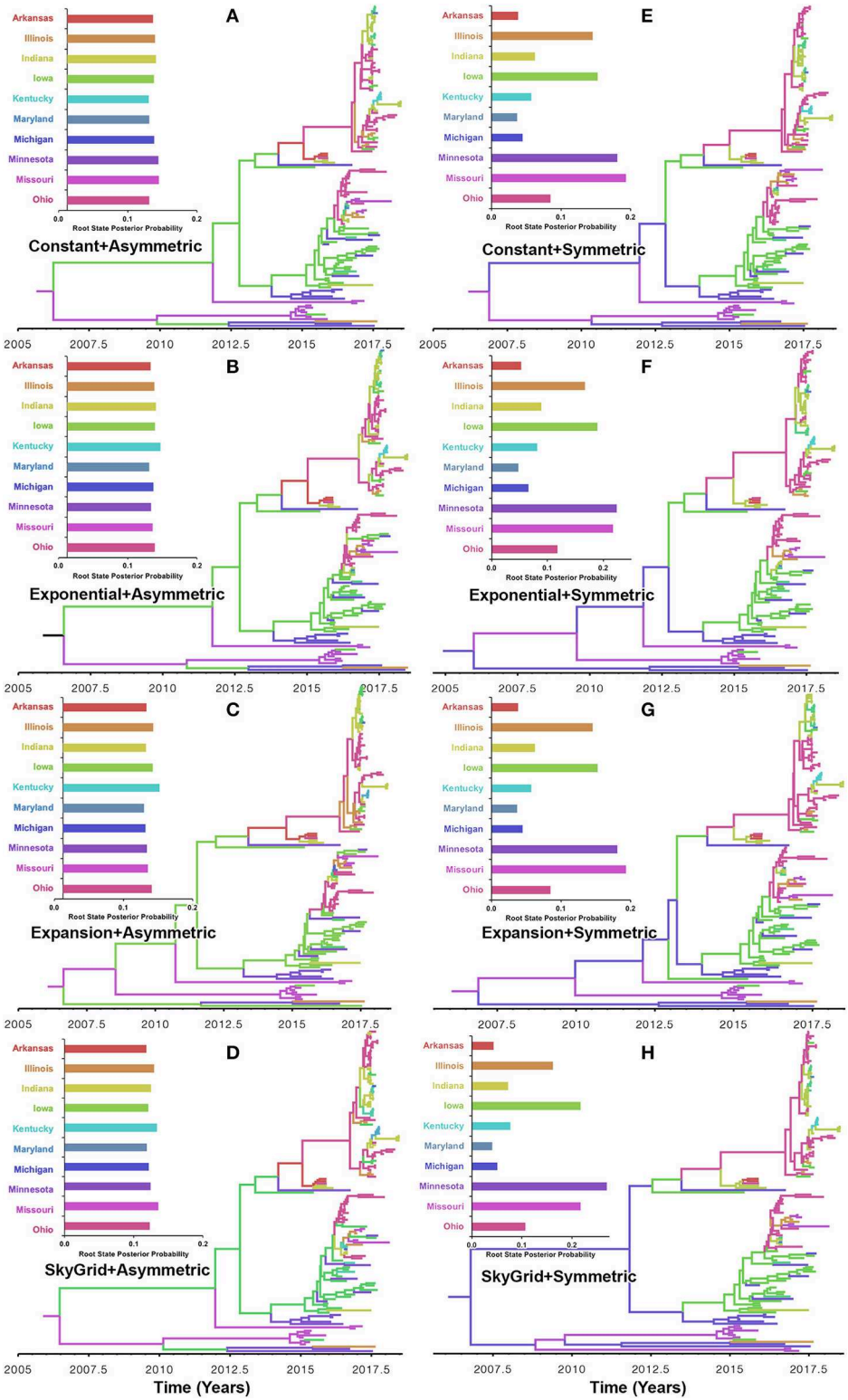


FIGURE 6 | Maximum clade credibility (MCC) phylogeny of the PB2 segment of human-like H3 swine influenza virus collected between January 2015 and June 2018 in the United States. The trees are inferred from eight phylodynamic model combinations (node-age and BSSVS priors). The color of the branches represents the most probable location state of their descendant nodes, and their color-coding corresponds to the upper left bar chart, which represents the root location state posterior probabilities (RSPP) for each state. **(A–D)** Trees inferred from four node-age + asymmetric BSSVS priors. **(E–H)** Trees inferred from four node-age + symmetric BSSVS priors.

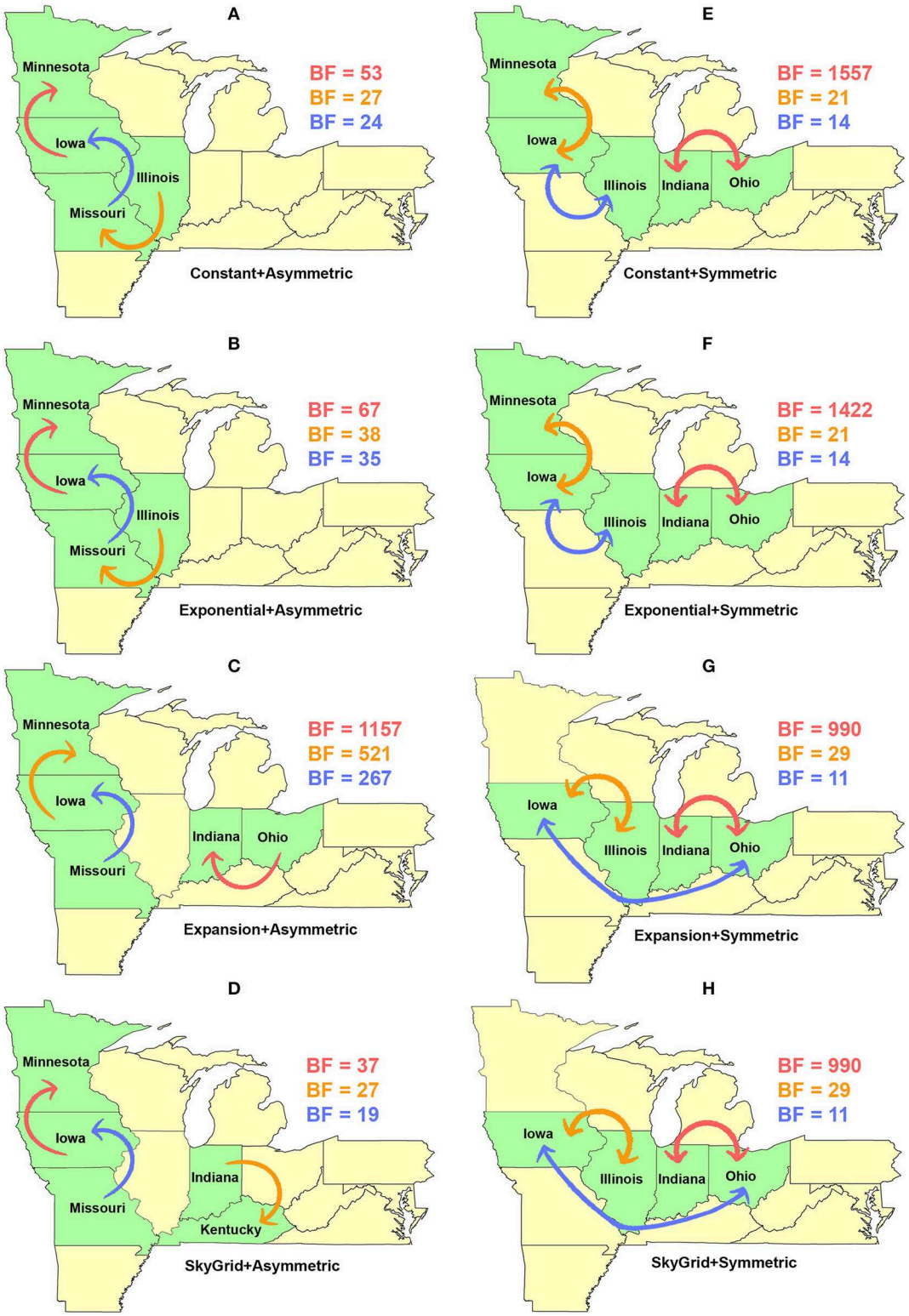


FIGURE 7 | Dispersal routes of human-like H3 swine influenza virus between states inferred from the HA gene segment. Dispersal routes with non-zero rates were inferred using the Bayesian stochastic search variable selection (BSSVS) approach, and statistically significant routes were selected using Bayes factors (BF). The top three dispersal routes with the strongest statistical support (by the BFs) are plotted. Arrows' colors correspond to the color legend of their BF values on the upper right of each map. **(A–D)** Dispersal routes inferred from four node-age + asymmetric BSSVS priors. **(E–H)** Dispersal routes inferred from four node-age + symmetric BSSVS priors.

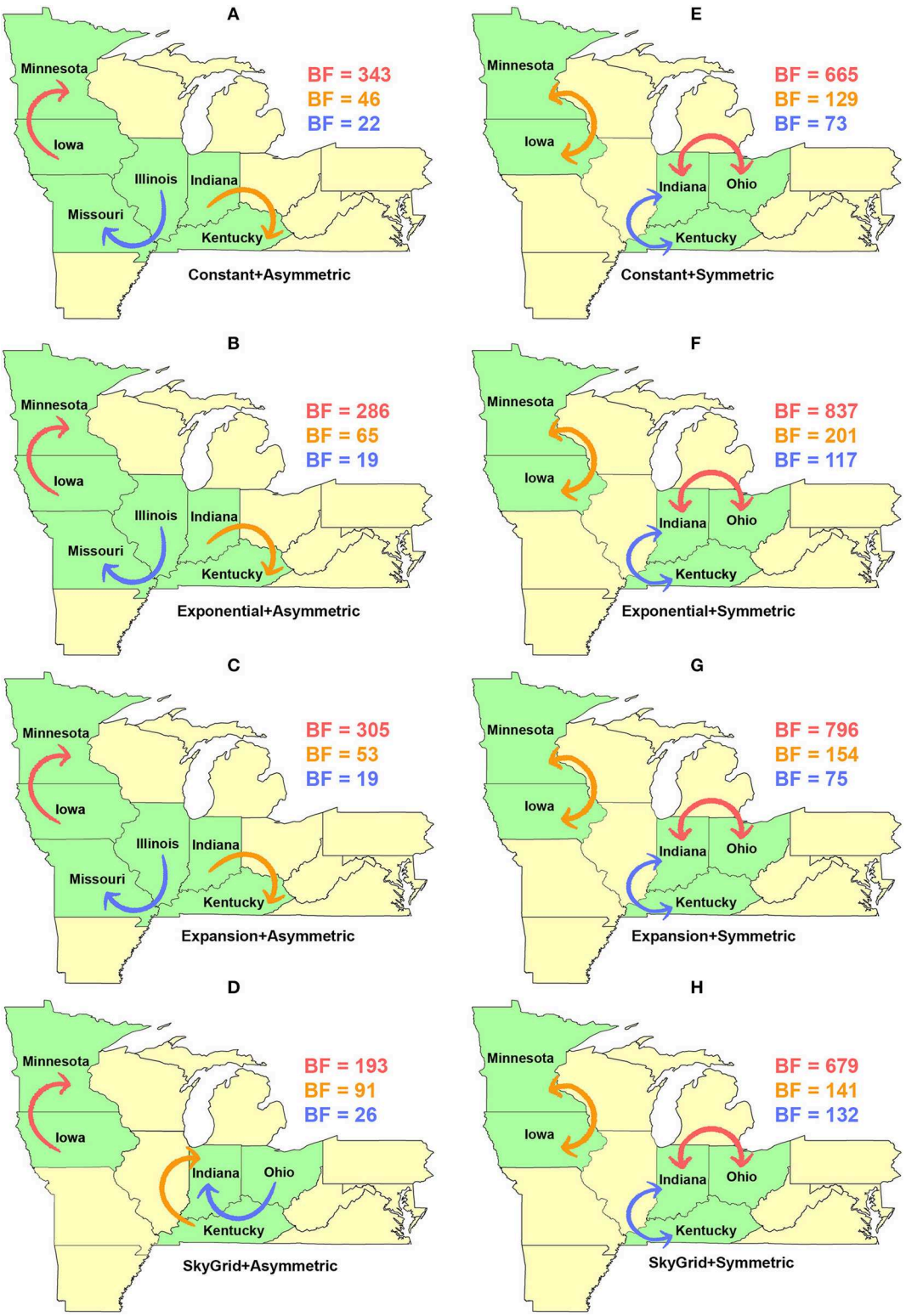


FIGURE 8 | Dispersal routes of human-like H3 swine influenza virus between states inferred from the PB2 gene segment. Dispersal routes with non-zero rates were inferred using the Bayesian stochastic search variable selection (BSSVS) approach, and statistically significant routes were selected using Bayes factors (BF). The top three dispersal routes with the strongest statistical support (by the BFs) are plotted. Arrows' colors correspond to the color legend of their BF values on the upper right of each map. **(A–D)** Dispersal routes inferred from four node-age + asymmetric BSSVS priors. **(E–H)** Dispersal routes inferred from four node-age + symmetric BSSVS priors.

(117). Furthermore, inferred relative genetic diversity through time did not decline for the HA segment and showed evidence of seasonal variation between 2014 and 2018 (**Figures 4A,B**), while a slight decline in the genetic diversity was inferred for the PB2 segment between 2015 and 2018 (**Figures 4C,D**). These findings suggest that currently circulating human-like H3 viruses will continue evolutionary activity leading to the generation of novel strains, which is attributed to the frequent and continuous exchange of viruses between commercial and exhibition swine operations in the United States with the later as the epicenter of that exchange (117). Our estimates of the TMRCAs for HA segment slightly agree on the notion that the oldest H3 viruses diverged from earlier outbreaks in the state of Minnesota, which is a central region for the swine industry in the United States (**Figure 3**). However, the notable overlap in the inferred 95% HPDs of the TMRCAs between most states (**Figure 3B**) suggest that the currently circulating strains are shifting their evolutionary dynamics in terms of re-emergence and dispersal when compared to earlier strains. Additionally, both gene segments agree on the assumption that H3 outbreaks were recently introduced into the state of Maryland (**Figure 3**).

The state of Minnesota was inferred to be the ancestral location of human-like H3 viruses isolated from outbreaks observed between 2009 and 2012 (118), which agrees with our TMRCAs inferred from HA segment (**Figure 3**). However, results of the SG + UCLN symmetric model, selected as the best fitting model for HA sequence data (**Supplementary Table 4**), implicates the state of Iowa as the ancestral region (after 2013) for currently circulating human-like H3 viruses, followed by the state of Minnesota as a secondary ancestral location (**Figure 5H**). This is not surprising since Iowa and Minnesota share the most prominent swine production system in the United States with the highest swine density, unrestricted and intense movement of animals between states. Although Iowa and Minnesota are the original hotspots of H3 viruses, our BSSVS BF results showed a markedly significant viral dispersal route between Indiana and Ohio (BF = 990) (**Figure 7H**). This suggests that the H3 viral gene flow between Ohio and Indiana, inferred for 2009–2012 viruses remains a vital migration route since, particularly within exhibition swine populations (117). Even though Illinois and Indiana formulate one swine production system, there was no significant viral dispersal route inferred between the states. Despite the continuous nature of animal movement within the production system of Minnesota and Iowa, no significant dispersal route was inferred between the two states using the HA segment (**Figure 7H**). Nevertheless, using the PB2 segment, a highly significant dispersal route was inferred from Iowa to Minnesota, suggesting that Iowa might be the new epicenter for virus dispersal of the currently circulating H3 lineages (**Figure 8D**). This result is further supported by the significant migration route between Iowa on one side and Illinois and Ohio on the other when using the HA segment (**Figure 7H**). Also, the inferred dispersal route between Iowa and Illinois (**Figure 7H**) may reflect interstate movements of exhibition pigs (119). Hence, the movements of exhibition pigs across the United States possibly led to expanding the spatial spread of H3 viruses to states with limited swine production systems (117).

Unlike the HA segment, RSPPs inferred from the most realistic phylodynamic model for PB2 sequences (i.e., asymmetric + SG + UCED) (**Supplementary Table 5**) did not yield conclusive results about the ancestral geographical origin of human-like H3 in the United States (**Figure 6D**). Instead, this result demonstrates a homogenous spatio-temporal diffusion process of the PB2 gene between states (**Figure 6D**), suggesting that the virus has maintained an endemic status across the United States after 2010. Also, results of the SG plot for PB2, described above, showed an overall stationarity in its genetic diversity through time (despite the slight early incline and later decline) (**Figure 4C**), when compared to the HA gene (**Figures 4A,B**), supporting the notion of endemic status. However, using the PB2 segment, we inferred a notably significant dispersal route originating from Iowa to Minnesota (BSSVS = 193) (**Figure 8D**), reflecting a well-established swine transportation route within a production system, as described above. However, this route was not inferred as significant when using the best fitting model for the HA segment (**Figure 7H**). These results may be attributed to the fact that PB2 evolutionary dynamics are moderately slower than the HA segment (**Figure 2**) in terms of strength of the temporal signal (**Supplementary Figure 2**), substitution rate (**Figure 2**), and age of the segment (**Figure 3**). Therefore, the PB2 segment maintained similar evolutionary dynamics to earlier strains that emerged in Minnesota and dispersed into Iowa (120). Yet, both HA and PB2 segment agree on the importance of Iowa as a geographical region for dispersal of currently circulating H3 lineages (**Figures 7H**, **8D**). Additionally, we inferred two significant viral dispersal routes originating from Kentucky to Indiana and from Ohio to Indiana (**Figure 8D**), which further supports the role of exhibition of swine movements between states in maintaining the spread of H3 viruses. Both dispersal routes are mainly maintained by the annual agricultural fairs where exhibition susceptible swine and humans from these states are frequently exposed to direct and indirect contacts from the same infected hosts (121). It is worth noting that the route from Kentucky to Indiana was hypothesized to be important for H3 gene flow between states, but past evolutionary analyses did not observe it due to the lack of sufficient samples (117).

Robustness and Limitations of Phylodynamic Methods

The uneven sampling of sequences in terms of temporal depth and frequency of associated discrete traits is an inherent limitation of most phylodynamic studies. For example, the inclusion of many recent sequences from a single geographical location may lead to a biased bottleneck effect in the shape of inferred population size through time when using a coalescent model from the Skyline family (122). This issue can be resolved by designing studies with uniform probability sampling with respect to space and time (122). Further, setting DTA is user friendly and computationally more efficient when compared to more complex coalescent models, but it underlays a few assumptions, such as that the sequence sample size is proportional to the size of the selected discrete state (26). Thus, including sequences

from severely undersampled discrete traits will tend to produce unreliable posterior inferences, where for example, inferred RSPPs will be skewed toward oversampled areas. Nevertheless, undersampling is a common problem, especially in passive surveillance data, and therefore, the use of structural coalescent models (e.g., BASTA) might be more appropriate (26).

Despite this inherited sensitivity of phylodynamic methods to uneven sampling, our posterior inference from the best fitting models showed remarkable robustness toward such limitation. Although the largest number of collected sequences was in 2017 (80) (**Supplementary Table 2**), estimates of relative genetic diversity through time did not show any striking jumps in that year for both HA and PB2 segments (**Figures 4B**, **2D**). Additionally, for the HA gene, Iowa (with 26 sequences) rather than Ohio (39 sequences) was inferred as the ancestral location (**Figure 5H**, **Supplementary Table 2**). However, seven out of the 17 U.S. states were excluded from the DTA due to the lack of sufficient sequences, and therefore, their role was unquantified in shaping the spatio-temporal evolution of SI. Yet, these states had substantially fewer swine-related activities as well as SI outbreaks than analyzed states.

Further, we showed how the posterior estimates of demographic reconstruction were almost insensitive to the choice of different phylodynamic priors for each gene segment (**Figures 2–4**). However, inferred evolutionary estimates from different gene regions may differ (41) or coincide (118) due to the natural variation in their mutation rate over time. This raises the question of whether using longer gene segments or whole genomes provides deeper resolution into the evolutionary history of rapidly evolving pathogens. Past influenza A studies (41, 123, 124), including the present study, showed that HA and NA segments typically exhibit higher evolutionary rates than more conserved segments like PB1 and PB2. Subsequently, segments with higher evolutionary rate will also display stronger evolutionary signals, as described above. In our analyses, the width of the 95% HPDs (i.e., length of the time scale) for the median age and TMRCAs of PB2 were remarkably wider than the HA segment (**Figures 2**, **3**). This sizeable width of the posterior intervals reflects the magnitude of uncertainty surrounding inferences from the PB2 segment, as well as suggests that inferences from the HA segment were more precise (or robust) than the PB2 segment. Also, we demonstrated how the PB2 segment failed to identify the ancestral geographical location of currently circulating H3 viruses (**Figure 6D**). While, using the symmetric model, we inferred four candidate ancestral locations with inconclusive RSPPs (**Figure 6H**). Further, Nelson et al. (117) were not able to infer a significant migration route between Indiana and Ohio using the PB2 segment. Yet, we were able to infer this particular route as significant using both the HA and the PB2 segments (**Figures 7H**, **8D**). Additionally, Scotch et al. (118) confirmed agreements in the phylogeographic inferences between HA and NA gene segments. This highlights another decisive question about the suitability and efficiency of using single, multiple, or whole genome when using phylodynamic methods for molecular surveillance of viral diseases. Most researchers advocate for whole-genome analysis by either analyzing each segment alone or as concinnated segments. However, in the presence of a large number of sequences, these strategies are ill timed and require massive computational resources, making them inefficient for targeted and near-real-time surveillance systems. It is worth noting that substitution rate and divergence time inferred by Alkhamis et al. (43) using the FMD SAT1 VP1 segment were similar to the evolutionary estimates inferred by Lasecka-Dykes et al. (125) using whole-genome sequences, confirming the robustness of phylodynamic methods. Nevertheless, the presence of recombination events can severely impact the robustness of phylodynamic methods leading to inferring biased evolutionary histories (126). Hence, targeting the most rapidly evolving gene segment at the beginning of an epidemic may suffice molecular surveillance activities. That said, the choice between gene segments or the whole genome should depend on the evolutionary properties of the pathogen, frequency of recombination events, availability of resources, and objectives of the molecular surveillance system.

As described above, phylodynamic inferences tend to be biased toward the available subsets of sequences data. Hence, when analyzing novel sequence datasets, it is critical to combine them with genetically related lineages published in the scientific literature or publicly available databases to reduce the impact of sampling bias as well as improve the reliability and accuracy of posterior evolutionary inferences. Unfortunately, several examples published in the scientific literature used phylodynamic methods on novel sequence datasets while ignoring their published relatives (127–129). This led to inferring MCC trees with unaccounted phylogenetic relationships such as nodes, branches, and roots.

Our worked example opens considerations for future work involving the use of more complex phylodynamic models, described above, to shed deeper insights into the evolutionary epidemiology of SI. For example, when the exact geographical locations of the sequences are available, the use of continuous phylogeographic models will enable us to include all states in the analyses, including states with few sequences. Besides, we can estimate the spatiotemporal dispersal speed of the virus as well as identify dispersal patterns (i.e., homogeneous vs. heterogeneous) across different geographical regions. Also, the use of GLM geographical models can directly quantify the importance of different environmental (e.g., climate) and demographical (e.g., pig density) factors in shaping the evolutionary history of SI in the United States. Finally, exploring the potentials of structured coalescent models in improving the reliability of inferences derived from basic DTAs should be considered as well.

Future of Phylodynamic Methods for Molecular Surveillance of Animal Diseases

The current surveillance programs rely heavily on collecting and analyzing spatial, temporal, and genomic aspects of an outbreak using classical statistical methods in a disjointed analytical framework. This disjointed framework suffers from many biases and is not capable of answering more profound epidemiological questions about the outbreak of current dynamics. Using our suggested phylodynamic analytical pipeline, we were able to fulfill critical epidemiological questions about

the emergence and evolution of currently circulating human-like H3 SI viruses in the United States, with the primary goal of guiding risk-based surveillance resources. For example, using inferences from the HA segment, we were able to identify the dates of epidemic introduction to each state. Also, we were able to identify the geographic origins of the current outbreaks and observed their genomic-spatio-temporal diffusion process through time between states. Also, we identified high-risk viral dispersal routes between states, rank-ordered their significance, and defined their directions. All of these are integral components of an effective risk-based molecular surveillance program, and the ability to achieve in real time is the future molecular surveillance of animal diseases. Nevertheless, the availability of computational resources for designing an ongoing phylodynamic-based molecular surveillance system will always remain a challenge, especially for developing countries. That said, a few open-source software developed recently can perform basic phylodynamic analysis (e.g., estimate molecular clocks and infer evolutionary models) using an ML statistical framework, including TimeTree (130), treedater R package (131), and Least Square Dating (120). While the algorithms implemented in these software trades off the advantages of the Bayesian framework, in the presence of large sequence datasets, they can produce evolutionary estimates similar to those estimated by BEAST using substantially less computational resources (120, 130, 131).

Nextstrain (https://nextstrain.org), which implements TreeTime is a futuristic working example of a web-based real-time molecular surveillance system for important human pathogens such as influenza, Ebola, Dengue, and the newly emerging corona (COVID-19) viruses. This surveillance system has an on-going phylodynamic analytical engine that traces, in real-time genetic diversity, divergence times, geographical origins, and dispersal on global scales. The system updates the results of the MCC tree once new sequences are deposited in other web-based publicly available genomic databases. However, this project is achieved through rigorous and consistent global collaboration and data sharing. In the United States, resources for developing a similar system for tracing animal diseases are readily available. Nevertheless, the chain of collaboration between researchers, government, and producers in the animal sector is hard to maintain due to logistic, economic, and educational (i.e., lack of awareness and skill in phylodynamic methods) reasons. Nevertheless, recent scientific literature on the use of phylodynamic methods for animal disease surveillance is notably growing, which reflects the increased awareness between veterinarians about the capacities of such methods and the goodwill of the industry leaders to voluntarily share their data (37, 132). Therefore, we anticipate a new era of animal disease prevention and control in the United States. In contrast, veterinary infrastructure in developing countries is severely lacking, in terms of reporting and data sharing, when compared to their human health sectors. Consequently, the question related to the future of implementing phylodynamic methods in global animal surveillance remains unanswered.

CONCLUSIONS

Our selected phylodynamic analytical pipeline offers an integrated approach to not only answering more profound epidemiological questions about emerging and endemic animal diseases but also attempts to reduce the impact of its inherent limitations to offer less biased and biologically plausible inferences about the pathogen evolutionary characteristics to properly guide intervention activities. This study has highlighted the value of phylodynamic methods in improving current and future molecular surveillance efforts against animal diseases using human-like H3 SI virus as a working example. We reviewed and outlined basic concepts and aspects of phylodynamic methods and attempted to summarize essential components of the methodology in one analytical pipeline to facilitate the proper use of the methods by animal health researchers. We also pinpointed requirements and challenges for integrating phylodynamic methods in routine animal disease surveillance activities.

DATA AVAILABILITY STATEMENT

The datasets analyzed for this study (Alignments, BEAST xmls, and MCC tree files) can be found in the Figshare Dataset. https://doi.org/10.6084/m9.figshare.11842989.v1.

AUTHOR CONTRIBUTIONS

The study was designed by MA and MT. The data were collected and organized by CL. All statistical analyses were conducted by MA. MA, CL, and MT wrote the first draft of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.00176/full#supplementary-material

REFERENCES

1. Perez A, AlKhamis M, Brito B, Carrasco-Medanic R, Whedbee Z, Willeberg P, et al. Carlsson: global animal disease surveillance. *Spat Spatiotemporal Epidemiol.* (2011) 2:135–45. doi: 10.1016/j.sste.2011.07.006
2. Alvarez J, Valdes-Donoso P, Tousignant S, Alkhamis M, Morrison R, Perez A. Novel analytic tools for the study of porcine reproductive and respiratory syndrome virus (PRRSv) in endemic settings: lessons learned in the U.S. *Porcine Health Manag.* (2016) 2:1–9. doi: 10.1186/s40813-016-0019-0
3. Vilalta C, Arruda AG, Tousignant SJP, Valdes-Donoso P, Muellner P, Muellner U, et al. A Review of quantitative tools used to assess the epidemiology of porcine reproductive and respiratory syndrome in U.S. swine farms using dr. morrison's swine health monitoring program data. *Front Vet Sci.* (2017) 4:94. doi: 10.3389/fvets.2017.00094
4. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet.* (2018) 19:9–20. doi: 10.1038/nrg.2017.88

5. Cattoli G, Monne I, Fusaro A, Joannis TM, Lombin LH, Aly MM, et al. Highly pathogenic avian influenza virus subtype H5N1 in Africa: a comprehensive phylogenetic analysis and molecular characterization of isolates. *PLoS ONE.* (2009) 4:e4842. doi: 10.1371/journal.pone.0004842
6. Walia RR, Anderson TK, Vincent AL. Regional patterns of genetic diversity in swine influenza a viruses in the United States from 2010 to 2016. *Influenza Other Respir Viruses.* (2018) 13:262–73. doi: 10.1111/irv.12559
7. Wallace RG, Fitch WM. Influenza aH5N1 immigration is filtered out at some international borders. *PLoS ONE.* (2008) 3:e1697. doi: 10.1371/journal.pone.0001697
8. Knowles NJ, Samuel AR. Molecular epidemiology of foot-and-mouth disease virus. *Virus Res.* (2003) 91:65–80. doi: 10.1016/S0168-1702(02)00260-5
9. Ahn I, Bae SE, Son HS. Comparative study of codon substitution patterns in foot-and-mouth disease virus (serotype O). *Exp Mol Med.* (2011) 43:587–95. doi: 10.3858/emm.2011.43.10.066
10. Jamal SM, Belsham GJ. Foot-and-mouth disease: past, present and future. *Vet Res.* (2013) 44:116. doi: 10.1186/1297-9716-44-116
11. Rosendal T, Dewey C, Friendship R, Wootton S, Young B, Poljak Z. Spatial and temporal patterns of porcine reproductive and respiratory syndrome virus (PRRSV) genotypes in Ontario, Canada, 2004-2007. *BMC Vet Res.* (2014) 10:83. doi: 10.1186/1746-6148-10-83
12. Martin-Valls GE, Kvisgaard LK, Tello M, Darwich L, Cortey M, Burgara-Estrella AJ, et al. Analysis of ORF5 and full-length genome sequences of porcine reproductive and respiratory syndrome virus isolates of genotypes 1 and 2 retrieved worldwide provides evidence that recombination is a common phenomenon and may produce mosaic isolates. *J Virol.* (2014) 88:3170–81. doi: 10.1128/JVI.02858-13
13. Stadejek T, Stankevicius A, Murtaugh MP, Oleksiewicz MB. Molecular evolution of PRRSV in Europe: current state of play. *Vet Microbiol.* (2013) 165:21–8. doi: 10.1016/j.vetmic.2013.02.029
14. Suchard MA, Weiss RE, Sinsheimer JS. Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol.* (2001) 18:1001–13. doi: 10.1093/oxfordjournals.molbev.a003872
15. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, et al. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proc Natl Acad Sci USA.* (2012) 109:15066–71. doi: 10.1073/pnas.1206598109
16. Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet.* (2009) 10:540–50. doi: 10.1038/nrg2583
17. Grubaugh ND, Ladner JT, Lemey P, Pybus OG, Rambaut A, Holmes EC, et al. Tracking virus outbreaks in the twenty-first century. *Nat Microbiol.* (2019) 4:10–19. doi: 10.1038/s41564-018-0296-2
18. Haydon DT, Bastos AD, Knowles NJ, Samuel AR. Evidence for positive selection in foot-and-mouth disease virus capsid genes from field isolates. *Genetics.* (2001) 157:7–15.
19. Lemey P, Rambaut A, Welch JJ, Suchard MA. Phylogeography takes a relaxed random walk in continuous space and time. *Mol Biol Evol.* (2010) 27:1877–85. doi: 10.1093/molbev/msq067
20. Magee D, Beard R, Suchard MA, Lemey P, Scotch M. Combining phylogeography and spatial epidemiology to uncover predictors of H5N1 influenza a virus diffusion. *Arch Virol.* (2015) 160:215–24. doi: 10.1007/s00705-014-2262-5
21. Alkhamis MA, Arruda AG, Morrison RB, Perez AM. Novel approaches for spatial and molecular surveillance of Porcine Reproductive and Respiratory Syndrome Virus (PRRSv) in the United States. *Sci Rep.* (2017) 7:4343. doi: 10.1038/s41598-017-04628-2
22. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, et al. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Pathog.* (2014) 10:e1003932. doi: 10.1371/journal.ppat.1003932
23. Nomikou K, Hughes J, Wash R, Kellam P, Breard E, Zientara S, et al. Widespread reassortment shapes the evolution and epidemiology of bluetongue virus following European invasion. *PLoS Pathog.* (2015) 11:e1005056. doi: 10.1371/journal.ppat.1005056
24. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* (2018) 4:vey016. doi: 10.1093/ve/vey016
25. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol.* (2012) 29:2157–67. doi: 10.1093/molbev/mss084
26. De Maio N, Wu CH, O'Reilly KM, Wilson D. New routes to phylogeography: a Bayesian structured coalescent approximation. *PLoS Genet.* (2015) 11:e1005421. doi: 10.1371/journal.pgen.1005421
27. Di Nardo A, Knowles NJ, Wadsworth J, Haydon DT, King DP. Phylodynamic reconstruction of O CATHAY topotype foot-and-mouth disease virus epidemics in the Philippines. *Vet Res.* (2014) 45:90. doi: 10.1186/s13567-014-0090-y
28. Su YCF, Bahl J, Joseph U, Butt KM, Peck HA, Koay ESC, et al. Phylodynamics of H1N1/2009 influenza reveals the transition from host adaptation to immune–driven selection. *Nat Commun.* (2015) 6:7952. doi: 10.1038/ncomms8952
29. Tian H, Feng Y, Vrancken B, Cazelles B, Tan H, Gill MS, et al. Transmission dynamics of re-emerging rabies in domestic dogs of rural China. *PLoS Pathog.* (2018) 14:e1007392. doi: 10.1371/journal.ppat.1007392
30. Brito B, Dee S, Wayne S, Alvarez J, Perez A. Genetic diversity of PRRS virus collected from air samples in four different regions of concentrated swine production during a high incidence season. *Viruses.* (2014) 6:4424–36. doi: 10.3390/v6114424
31. Shi M, Lemey P, Singh Brar M, Suchard MA, Murtaugh MP, Carman S, et al. The spread of type 2 Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) in North America: a phylogeographic approach. *Virology.* (2013) 447:146–54. doi: 10.1016/j.virol.2013.08.028
32. Magee D, Suchard MA, Scotch M. Bayesian phylogeography of influenza A/H3N2 for the 2014-15 season in the United States using three frameworks of ancestral state reconstruction. *PLoS Comput Biol.* (2017) 13:e1005389. doi: 10.1371/journal.pcbi.1005389
33. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol Biol Evol.* (2013) 30:713–24. doi: 10.1093/molbev/mss265
34. Hedge J, Lycett SJ, Rambaut A. Real-time characterization of the molecular epidemiology of an influenza pandemic. *Biol Lett.* (2013) 9:20130331. doi: 10.1098/rsbl.2013.0331
35. Lu L, Lycett SJ, Leigh Brown AJ. Reassortment patterns of avian influenza virus internal segments among different subtypes. *BMC Evol Biol.* (2014) 14:16. doi: 10.1186/1471-2148-14-16
36. Fourment M, Holmes EC. Avian influenza virus exhibits distinct evolutionary dynamics in wild birds and poultry. *BMC Evol Biol.* (2015) 15:120. doi: 10.1186/s12862-015-0410-5
37. Alkhamis MA, Perez AM, Murtaugh MP, Wang X, Morrison RB. Applications of bayesian phylodynamic methods in a recent porcine reproductive US. and respiratory syndrome virus outbreak. *Front Microbiol.* (2016) 7:67. doi: 10.3389/fmicb.2016.00067
38. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. HIV epidemiology. *t*he early spread and epidemic ignition of HIV−1 in human populations. *Science.* (2014) 346:56–61. doi: 10.1126/science.1256739
39. Zehender G, Pariani E, Piralla A, Lai A, Gabanelli E, Ranghiero A, et al. Reconstruction of the evolutionary dynamics of the A(H1N1)pdm09 influenza virus in Italy during the pandemic and post-pandemic phases. *PLoS ONE.* (2012) 7:e47517. doi: 10.1371/journal.pone.0047517
40. A.a.P.H. I. Service. *Influenza A Virus in Swine Surveillance*. In: *Fiscal Year*. (2018). Quarterly Report. Ed U.S. D. o. Agriculture (2018).
41. Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, et al. Rambaut: Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature.* (2009) 459:1122–5. doi: 10.1038/nature08182
42. Yoon SH, Park W, King DP, Kim H. Phylogenomics and molecular evolution of foot-and-mouth disease virus. *Mol Cells.* (2011) 31:413–21. doi: 10.1007/s10059-011-0249-6
43. Omondi G, Alkhamis MA, Obanda V, Gakuya F, Sangula A, Pauszek S, et al. Phylogeographic and cross-species transmission dynamics of SAT1 and SAT2 foot-and-mouth disease virus in eastern Africa. *Mol Ecol.* (2019) 28:2903–16. doi: 10.1111/mec.15125
44. Ekong PS, Fountain-Jones NM, Alkhamis MA. Spatiotemporal evolutionary epidemiology of H5N1 highly pathogenic avian influenza in West Africa and Nigeria, 2006–2015. *Transbound Emerg Dis.* (2018) 65:e70–82. doi: 10.1111/tbed.12680
45. Bahl J, Krauss S, Kuhnert D, Fourment M, Raven G, Pryor SP, et al. Influenza a virus migration and persistence in north american wild birds. *PLoS Pathog.* (2013) 9:e1003570. doi: 10.1371/journal.ppat.1003570

46. Wong KM, Suchard MA, Huelsenbeck JP. Alignment uncertainty and genomic analysis. *Science.* (2008) 319:473–6. doi: 10.1126/science.1151532
47. Ashkenazy H, Sela I, Levy Karin E, Landan G, Pupko T. Multiple sequence alignment averaging improves phylogeny reconstruction. *Syst Biol.* (2019) 68:117–30. doi: 10.1093/sysbio/syy036
48. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and clustal X version 2.0. *Bioinformatics.* (2007) 23:2947–8. doi: 10.1093/bioinformatics/btm404
49. Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* (2000) 302:205–17. doi: 10.1006/jmbi.2000.4042
50. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* (2004) 32:1792–7. doi: 10.1093/nar/gkh340
51. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics.* (2014) 30:3276–8. doi: 10.1093/bioinformatics/btu531
52. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol.* (1981) 17:368–76. doi: 10.1007/BF01734359
53. Farris JS. Methods for computing wagner trees. *Sys Zool.* (1970) 19:83–92. doi: 10.2307/2412028
54. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* (1987) 4:406–25.
55. Hall BG. Building phylogenetic trees from molecular data with MEGA. *Mol Biol Evol.* (2013) 30:1229–35. doi: 10.1093/molbev/mst012
56. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* (2014) 30:1312–3. doi: 10.1093/bioinformatics/btu033
57. Holmes EC. The phylogeography of human viruses. *Mol Ecol.* (2004) 13:745–56. doi: 10.1046/j.1365-294X.2003.02051.x
58. Menardo F, Loiseau C, Brites D, Coscolla M, Gygli SM, Rutaihwa LK, et al. Treemmer: a tool to reduce large phylogenetic datasets with minimal loss of diversity. *BMC Bioinformatics.* (2018) 19:164. doi: 10.1186/s12859-018-2164-8
59. Perez-Losada M, Arenas M, Galan JC, Palero F, Gonzalez-Candelas F. Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. *Infect Genet Evol.* (2015) 30:296–307. doi: 10.1016/j.meegid.2014.12.022
60. Schierup MH, Hein J. Consequences of recombination on traditional phylogenetic analysis. *Genetics.* (2000) 156:879–91.
61. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics.* (2010) 26:2462–3. doi: 10.1093/bioinformatics/btq467
62. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using tempEst (formerly path-O-Gen). *Virus Evol.* (2016) 2:vew007. doi: 10.1093/ve/vew007
63. Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG. Measurably evolving populations. *Trends Ecol Evol.* (2003) 18:481–8. doi: 10.1016/S0169-5347(03)00216-7
64. Drummond AJ, Bouckaert R. Substitution and site models. In: *Bayesian Evolutionary Analysis with BEAST*. Cambridge: Cambridge University Press (2015) 44–57.
65. Minin VN, Suchard MA. Fast, accurate and simulation-free stochastic mapping. *Philos Trans R Soc Lond B Biol Sci.* (2008) 363:3985–95. doi: 10.1098/rstb.2008.0176
66. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* (1985) 22:160–74. doi: 10.1007/BF02101694
67. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures On Mathematics In The Life Sci.* (1986) 17:57–86.
68. Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol.* (1994) 39:306–14 doi: 10.1007/BF00160154
69. Gu X, Fu YX, Li WH. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol Biol Evol.* (1995) 12:546–57.
70. Ren F, Tanaka H, Yang Z. An empirical examination of the utility of codon-substitution models in phylogeny reconstruction. *Syst Biol.* (2005) 54:808–18. doi: 10.1080/10635150500354688
71. Yang Z. Maximum-likelihood models for combined analyses of multiple sequence data. *J Mol Evol.* (1996) 42:587–96 doi: 10.1007/BF02352289
72. Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol.* (1998) 15:568–73. doi: 10.1093/oxfordjournals.molbev.a025957
73. Schwarz G. Estimating the dimension of a model. *Ann Stat.* (1978) 6: 461–4. doi: 10.1214/aos/1176344136
74. Hurvich CM, Tsai CL. Regression and time series model selection in small samples. *Biometrika.* (1989) 76:297–307. doi: 10.1093/biomet/76.2.297
75. Bao L, Gu H, Dunn KA, Bielawski JP. Methods for selecting fixed-effect models for heterogeneous codon evolution, with comments on their application to gene and genome data. *BMC Evol Biol.* (2007) 7 (Suppl. 1):S5. doi: 10.1186/1471-2148-7-S1-S5
76. Lanfear R, Calcott B, Ho SY, Guindon S. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol.* (2012) 29:1695–701. doi: 10.1093/molbev/mss020
77. Darriba D, Taboada GL, Doallo R, Posada D. ModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* (2012) 9:772. doi: 10.1038/nmeth.2109
78. Bouckaert RR, Drummond AJ. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evol Biol.* (2017) 17:42. doi: 10.1186/s12862-017-0890-6
79. Drummond AJ, Bouckaert R. The molecular clock. In *Bayesian Evolutionary Analysis With BEAST*. Cambridge: Cambridge University Press (2015) 58–67.
80. Thorne JL, Kishino H. Divergence time and evolutionary rate estimation with multilocus data. *Syst Biol.* (2002) 51:689–702. doi: 10.1080/10635150290102456
81. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* (2006) 4:e88. doi: 10.1371/journal.pbio.0040088
82. Yang Z, Yoder AD. Estimation of the transition/transversion rate bias and species sampling. *J Mol Evol.* (1999) 48:274–83. doi: 10.1007/PL00006470
83. Drummond AJ, Suchard MA. Bayesian random local clocks, or one rate to rule them all. *BMC Biol.* (2010) 8:114. doi: 10.1186/1741-7007-8-114
84. Drummond AJ, Bouckaert R. Evolutionary trees. In *Bayesian Evolutionary Analysis with BEAST*. Cambridge: Cambridge University Press (2015) 84:21–43.
85. Kingman JFC. On the genealogy of large populations. *J Appl Probab.* (1982) 19:27–43. doi: 10.2307/3213548
86. Kuhner MK, Yamato J, Felsenstein J. Estimating effective population size and mutation rate from sequence data using Metropolis–hastings sampling. *Genetics.* (1995) 140:1421–30.
87. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics.* (2002) 161:1307–20.
88. Beerli P, Felsenstein J. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics.* (1999) 152:763–73.
89. Bloomquist EW, Suchard MA. Unifying vertical and nonvertical evolution: a stochastic ARG-based framework. *Syst Biol.* (2010) 59:27–41. doi: 10.1093/sysbio/syp076
90. Seo TK, Thorne JL, Hasegawa M, Kishino H. A viral sampling design for testing the molecular clock and for estimating evolutionary rates and divergence times. *Bioinformatics.* (2002) 18:115–23. doi: 10.1093/bioinformatics/18.1.115
91. Griffiths RC, Tavare S. Ancestral inference in population genetics. *Stat Sci.* (1994) 307–19 doi: 10.1214/ss/1177010378
92. Alkhamis MA, Moore BR, Perez AM. Phylodynamics of H5N1 highly pathogenic avian influenza in europe, 2005-2010: potential for molecular surveillance of new outbreaks. *Viruses.* (2015) 7:3310–28. doi: 10.3390/v7062773
93. Pybus OG, Rambaut A, Harvey PH. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics.* (2000) 155:1429–37.
94. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol.* (2005) 22:1185–92. doi: 10.1093/molbev/msi103

95. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: bayesian coalescent-based inference of population dynamics. *Mol Biol Evol.* (2008) 25:1459–71. doi: 10.1093/molbev/msn090
96. Drummond AJ, Bouckaert R. Structured trees and phylogeography. In: *Bayesian Evolutionary Analysis With BEAST*. Cambridge: Cambridge University Press (2015) 68–76.
97. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. *PLoS Comput Biol.* (2009) 5:e1000520. doi: 10.1371/journal.pcbi.1000520
98. Alkhamis MA, Gallardo C, Jurado C, Soler A, Arias M, Sanchez-Vizcaino JM. Phylodynamics and evolutionary epidemiology of african swine fever p72-CVR genes in Eurasia and Africa. *PLoS ONE.* (2018) 13:e0192565. doi: 10.1371/journal.pone.0192565
99. Ewing G, Nicholls G, Rodrigo A. Using temporally spaced sequences to simultaneously estimate migration rates, mutation rate and population sizes in measurably evolving populations. *Genetics.* (2004) 168:2407–20. doi: 10.1534/genetics.104.030411
100. De Maio N, C.-Wu H, Wilson DJ. SCOTTI: efficient reconstruction of transmission within outbreaks with the structured coalescent. *PLoS Comput Biol.* (2016) 12:e1005130. doi: 10.1371/journal.pcbi.1005130
101. Müller NF, Rasmussen D, Stadler T. MASCOT: parameter and state inference under the marginal structured coalescent approximation. *Bioinformatics.* (2018) 34:3843–8. doi: 10.1093/bioinformatics/bty406
102. Volz EM, Kosakovsky Pond SL, Ward MJ, Leigh Brown AJ, Frost SD. Phylodynamics of infectious disease epidemics. *Genetics.* (2009) 183:1421–30. doi: 10.1534/genetics.109.106021
103. Frost SD, Volz EM. Modelling tree shape and structure in viral phylodynamics. *Philos Trans R Soc Lond B Biol Sci.* (2013) 368:1614. doi: 10.1098/rstb.2012.0208
104. Leventhal GE, Günthard HF, Bonhoeffer S, Stadler T. Using an epidemiological model for phylogenetic inference reveals density dependence in HIV transmission. *Mol Biol Evol.* (2014) 31:6–17. doi: 10.1093/molbev/mst172
105. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* (2012) 29:1969–73. doi: 10.1093/molbev/mss075
106. BEAST developers. *Model Selection Tutorial.*(2019). Available online at: https://beast.community/model_selection_1. (accessed date: Janaury 30, 2020).
107. BEAST developers. *Model Selection Tutorial.*(2019). Available online at: https://beast.community/model_selection_2. (accessed date: Janaury 30, 2020).
108. BEAST developers (2019). *Model Selection Tutorial*. Available online at: https://beast.community/workshop_model_selection (accessed Janaury 30, 2020).
109. Fan Y, Wu R, Chen M-H, Kuo L, Lewis PO. Choosing among partition models in Bayesian Phylogenetics. *Mol Biol Evol.* (2011) 28:523–32. doi: 10.1093/molbev/msq224
110. Drummond AJ, Bouckaert R. Posterior analysis and post processing. In: *Bayesian Evolutionary Analysis with BEAST*. Cambridge: Cambridge University Press (2015) 139–152.
111. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Sys Biol.* (2018) 67:901–4. doi: 10.1093/sysbio/syy032
112. Steel M. Should phylogenetic models be trying to "fit an elephant"? *Trends Genet.* (2005) 21:307–9. doi: 10.1016/j.tig.2005.04.001
113. Rambaut: FigTree. Available online at: http://tree.bio.ed.ac.uk/software/figtree/. (2018).
114. Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P. SpreaD3: interactive visualization of spatiotemporal history and trait evolutionary processes. *Mol Biol Evol.* (2016) 33:2167–9. doi: 10.1093/molbev/msw082
115. Rajao DS, Gauger PC, Anderson TK, Lewis NS, Abente EJ, Killian ML, et al. Novel reassortant human-Like H3N2 and H3N1 influenza A Viruses detected in pigs are virulent and antigenically distinct from Swine viruses endemic to the United States. *J Virol.* (2015) 89:11213–22. doi: 10.1128/JVI.01675-15
116. Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, et al. Lipman: the influenza virus resource at the national center for biotechnology information. *J Virol.* (2008) 82:596. doi: 10.1128/JVI.02005-07
117. Nelson MI, Wentworth DE, Das SR, Sreevatsan S, Killian ML, Nolting JM, et al.S. Bowman: evolutionary dynamics of influenza A viruses in US exhibition swine. *J Infect Dis.* (2016) 213:173–82. doi: 10.1093/infdis/jiv399
118. Scotch M, Mei C. Phylogeography of swine influenza H3N2 in the United States: translational public health for zoonotic disease surveillance. *Infect Genet Evol.* (2013) 13:224–9. doi: 10.1016/j.meegid.2012.09.015
119. Nelson MI, Lemey P, Tan Y, Vincent A, Lam TT, Detmer S, et al. Spatial dynamics of human-origin H1 influenza a virus in North American swine. *PLoS Pathog.* (2011) 7:e1002077. doi: 10.1371/journal.ppat.1002077
120. T.-To H, Jung M, Lycett S, Gascuel O. Fast dating using least-Squares criteria and algorithms. *Sys Biol.* (2016) 65:82–97. doi: 10.1093/sysbio/syv068
121. Bliss N, Nelson SW, Nolting JM, Bowman AS. Prevalence of influenza a virus in exhibition swine during arrival at agricultural fairs. *Zoonoses Public Health.* (2016) 63:477–85. doi: 10.1111/zph.12252
122. Hall MD, Woolhouse ME J, Rambaut A. The effects of sampling strategy on the quality of reconstruction of viral population dynamics using Bayesian skyline family coalescent methods: a simulation study. *Virus Evol.* (2016) 2:vew003. doi: 10.1093/ve/vew003
123. Rahnama L, Aris-Brosou S. Phylodynamics of the emergence of influenza viruses after cross-species transmission. *PLoS ONE.* (2013) 8:e82486. doi: 10.1371/journal.pone.0082486
124. Bhatt S, Lam TT, Lycett SJ, Leigh Brown AJ, Bowden TA, Holmes EC, et al. The evolutionary dynamics of influenza a virus adaptation to mammalian hosts. *Philos Trans R Soc Lond B Biol Sci.* (2013) 368:20120382.
125. Lasecka-Dykes L, Wright CF, Di Nardo A, Logan G, Mioulet V, Jackson T, et al. Full genome sequencing reveals New Southern African Territories genotypes Bringing us closer to understanding true variability of foot-and-mouth disease virus in Africa. *Viruses.* (2018) 10:E192. doi: 10.3390/v10040192
126. Bachanek-Bankowska K, di Nardo A, Wadsworth J, Mioulet V, Pezzoni G, Grazioli S, et al. Reconstructing the evolutionary history of pandemic foot-and-mouth disease viruses: the impact of recombination within the emerging O/ME-SA/Ind-2001 lineage. *Sci Rep.* (2018) 8:14693. doi: 10.1093/ve/vez002.046
127. Boyle DB, Amos-Ritchie R, Broz I, Walker PJ, Melville L, Flanagan D, et al. Evolution of bluetongue virus serotype 1 in northern Australia over 30 years. *J Virol.* (2014) 88:13981–9. doi: 10.1128/JVI.02055-14
128. Lam TT, Hon CC, Lemey P, Pybus OG, Shi M, Tun HM, et al. Phylodynamics of H5N1 avian influenza virus in Indonesia. *Mol Ecol.* (2012) 21:3062–77. doi: 10.1111/j.1365-294X.2012.05577.x
129. Chaikhumwang P, Tantituvanont A, Tripipat T, Tipsombatboon P, Piriyapongsa J, Nilubol D. Dynamics and evolution of highly pathogenic porcine reproductive and respiratory syndrome virus following its introduction into a herd concurrently infected with both types 1 and 2. *Infect Genet Evol.* (2015) 30:64–74. doi: 10.1016/j.meegid.2014.12.025
130. Sagulenko P, Puller V, Neher RA. TreeTime: maximum-likelihood phylodynamic analysis. *Virus Evol.* (2018) 4:vex042. doi: 10.1093/ve/vex042
131. Volz EM, Frost SDW. Scalable relaxed clock phylogenetic dating. *Virus Evol.* (2017) 3:vex025. doi: 10.1093/ve/vex025
132. Rife BD, Mavian C, Chen X, Ciccozzi M, Salemi M, Min J, et al. Phylodynamic applications in 21(st) century global infectious disease research. *Glob Health Res Policy.* (2017) 2:13. doi: 10.1186/s41256-017-0034-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

METHODS
published: 19 May 2020
doi: 10.3389/fvets.2020.00271

How to Conduct a Bayesian Network Meta-Analysis

Dapeng Hu[1], *Annette M. O'Connor*[2]*, *Chong Wang*[1,3], *Jan M. Sargeant*[4] and *Charlotte B. Winder*[4]

[1] Department of Statistics, Iowa State University, Ames, IA, United States, [2] Department of Large Animal Clinical Sciences, College of Veterinary Medicine, Michigan State University, East Lansing, MI, United States, [3] Department of Veterinary Diagnostic and Production Animal Medicine, 2203 Lloyd Veterinary Medical Center, Iowa State University, Ames, IA, United States, [4] Department of Population Medicine, University of Guelph, Guelph, ON, Canada

Network meta-analysis is a general approach to integrate the results of multiple studies in which multiple treatments are compared, often in a pairwise manner. In this tutorial, we illustrate the procedures for conducting a network meta-analysis for binary outcomes data in the Bayesian framework using example data. Our goal is to describe the workflow of such an analysis and to explain how to generate informative results such as ranking plots and treatment risk posterior distribution plots. The R code used to conduct a network meta-analysis in the Bayesian setting is provided at GitHub.

Keywords: network meta-analysis, Bayesian, systematic review, tutorial, veterinary science

OPEN ACCESS

Edited by:
Andres M. Perez, University of Minnesota Twin Cities, United States

Reviewed by:
Fraser Lewis, University of Zurich, Switzerland
Gilles Kratzer, University of Zurich, in collaboration with reviewer FL
Adam Branscum, Oregon State University, United States

***Correspondence:**
Annette M. O'Connor
oconn445@msu.edu

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *11 December 2019*
Accepted: *22 April 2020*
Published: *19 May 2020*

Citation:
Hu D, O'Connor AM, Wang C, Sargeant JM and Winder CB (2020) How to Conduct a Bayesian Network Meta-Analysis. Front. Vet. Sci. 7:271. doi: 10.3389/fvets.2020.00271

1. INTRODUCTION

Meta-analysis is a quantitative method commonly used to combine the results of multiple studies in the medical and veterinary sciences. There are several common types of meta-analysis. A pairwise meta-analysis compares two treatments across multiple studies, whereas a network meta-analysis involves the simultaneous synthesis of multiple studies to create pairwise comparisons of more than two treatments. A third type of meta-analysis is multivariate meta-analysis, which is far less common than the other two types (1, 2). Regardless of the type, meta-analyses can be conducted using study-level summary data, which are usually reported in the literature. In the human health sciences, it is also possible to perform meta-analyses using data from individual patients, but meta-analysis using individual-level data is very rare in veterinary science (3).

In this tutorial, we focus on network meta-analysis, which is becoming increasingly common in both human health and the veterinary sciences (4–10). Although frequently used as a synonym for network meta-analysis, a mixed treatment comparisons meta-analysis is a type of network meta-analysis that can be described as a "*A statistical approach used to analyze a network of evidence with more than two interventions which are being compared indirectly, and at least one pair of interventions compared both directly and indirectly*" (1). Direct comparisons of interventions are obtained from trials or observational studies that include both interventions and compare them directly. Indirect comparisons of interventions, on the other hand, are made based on multiple trials that each included one, but not both, of the interventions of interest and therefore did not compare the interventions directly as part of the original study. In general, network meta-analysis offers the advantage of enabling the combined assessment of more than two treatments. A network meta-analysis that includes the mixed treatment comparisons "component" has the additional feature of enabling a formal statistical estimation of indirect treatment comparisons that might not be available in the literature (4, 7). Most network meta-analyses include a mixed treatment comparisons component, so we use the term network meta-analysis to refer to mixed treatment comparisons meta-analyses throughout this manuscript. There are some R (11) packages available for conducting Bayesian network meta-analysis such as *gemtc* (12) and *BUGSnet* (13). The output given by *gemtc* is limited. For example, *gemtc* does not have the option to report the summary effect

as either the relative risk or absolute risk. Further, the output is not available in a table format. While *BUGSnet* is limited to analyzing arm-level data which could be a limitation for veterinary data which is often reported at the contrast level.

1.1. Rationale

Currently, only a few systematic reviews in veterinary science have employed network meta-analysis. However, if the trend in the human health sciences is indicative of what will occur in veterinary science, we can expect to see more network meta-analyses of veterinary studies in the future. For example, in 2010, a PubMed search with the terms "network meta-analysis" OR "mixed treatment comparison" yielded 10 citations, whereas by 2018, the same search returned 618 citations. The rise in the use of network meta-analysis is a function of the value that such an analysis provides to the decision-making community. Instead of limiting comparisons to those that are made across just two interventions and published in the literature, as is the case for pairwise meta-analysis, network meta-analysis allows the simultaneous comparison of multiple treatments, including comparisons that are not directly available in the literature. For many clinical decisions in veterinary medicine, there are multiple interventions that could be used to prevent or treat a specific disease or condition. Therefore, decision-makers are interested in the comparative efficacy of all the options rather than just pairwise comparisons. To illustrate the limitations of pairwise meta-analysis, we can use the choice of which antibiotic to use to treat bovine respiratory disease as an example. The vast majority of publicly available trials involving antibiotics for bovine respiratory disease were conducted in order to register and license a particular product. In those types of trials, the antibiotic of interest is typically compared with a placebo to demonstrate that the antibiotic has a significant beneficial effect. Veterinarians are actually interested in the comparative efficacy of all the available antibiotics, but for a variety of reasons (e.g., economic, marketing, and regulatory), few head-to-head comparisons of antibiotics are available. A network meta-analysis can fill that information gap for veterinarians by providing head-to-head estimates of the comparative efficacy of antibiotics, even though those comparisons are not available in the literature.

1.2. Objectives

Our objective is to provide a tutorial illustrating how to conduct a network meta-analysis of study-level results from multiple sources. Network meta-analysis can be conducted using a frequentist approach or a Bayesian approach. We focus on the Bayesian approach for three reasons:

- First, Bayesian approaches to network meta-analysis are currently more common than frequentist approaches (14–16).
- Second, the learning curve for the Bayesian approach is steeper than that for the frequentist approach. There are several standard packages that can be used to conduct a frequentist analysis, and the examples provided with the packages are usually sufficient to enable the analysis to be conducted (17, 18). Therefore a tutorial for the Bayesian approach fills a larger gap.
- Third, the Bayesian approach allows for many outputs that enhance understanding of the data. For example, the point estimate, as well as the posterior distribution of the absolute risk of each treatment can be obtained from the results of the Bayesian approach. Therefore, a tutorial focused on the Bayesian approach to network meta-analysis has greater utility.

1.3. Target Audience

We describe the step-wise workflow of a network meta-analysis, and we provide R, JAGS (19) and BUGS (20) code for end-users interested in troubleshooting or optimizing their own analyses (see **Appendix** for link). It is not our intention to teach the statistical foundations of network meta-analysis. We believe that this tutorial will fill a gap between papers that explain the underlying statistical methodology and the "black box" tutorials that typically come with statistical packages. Our tutorial is intended for readers interested in understanding the software-coding and data-management processes that underlie a network meta-analysis. It is our hope that by using our tutorial, a reader would be able to find errors in his or her own network meta-analysis or modify existing code to produce a new output. We assume that the reader is familiar with pairwise meta-analysis [see the companion paper in the frontiers series (21) and the paper about synthesizing data from intervention studies using meta-analysis (22) for more details].

2. ORGANIZATION

The tutorial is organized in three parts. First, we provide a basic introduction to Bayesian network meta-analysis and the concepts in the underlying model. Second, we discuss how to conduct the analysis, with a focus on the software processes involved. Third (in the **Appendix**), we provide actual code that can be used to conduct a Bayesian network meta-analysis. The Appendix contains detailed instructions on how to run the R code that will perform the analysis and produce the desired outputs. The code includes R and jags scripts for executing a network meta-analysis in an R project, which contains several scripts that the reader can run to better understand the processes associated with conducting the analysis and obtaining the output. Not all readers will want to delve into the mechanisms of the Appendix code. For readers who want to conduct a network meta-analysis but are not interested in the mechanics of coding the analysis, we suggest that they read the first two parts of the tutorial and then use an R package that includes functions for running a network meta-analysis, such as *gemtc*.

3. THE BASICS OF NETWORK META-ANALYSIS

3.1. Arm-Level Data and Contrast-Level Data

The first part of a network meta-analysis is data extraction from the primary sources, preferably based on a systematic review conducted using an *a priori* protocol. The data extracted from

the primary sources are study-level summary data (also called aggregated data) in one of two forms: arm-level data, which report the effect measures (i.e., absolute odds or absolute risk) for each arm, or contrast-level data, which show the contrast of effects, or the effect size (23), between treatment arms (i.e., the odds ratio, relative risk, or log odds ratio). Either type of summary data can be used in a network meta-analysis using either the Bayesian or the frequentist approach (7).

It is essential, however, that the data extracted for a network meta-analysis meet the transitivity assumption, that is, that each enrolled subject in a given study would be eligible for enrollment in the other studies. For example, in a previous network meta-analysis of antibiotic treatments for bovine respiratory disease, data from studies that included antibiotic metaphylaxis were excluded, because animals that received prior antibiotic treatment would have limited eligibility for subsequent antibiotic treatments and would therefore violate the transitivity assumption (5, 6). Animals that received an antibiotic as a metaphylactic treatment would be unlikely to receive the same antibiotic as the first treatment of choice once bovine respiratory disease was diagnosed. Moreover, the effect of an antibiotic might be different if the antibiotic was previously used for metaphylactic treatment in the same animal, so the results from studies with and without metaphylaxis would not be the same. By limiting the network of eligible studies for the meta-analysis to those without metaphylaxis, the transitivity assumption would be more likely to hold.

3.2. The Comparative Effects Model

A key aspect of network meta-analysis is the comparative effects model. The comparative effects model forms the basis for the estimation of the relative treatment effects, which make up the main output of the network meta-analysis. A commonly used approach to network meta-analysis is to directly describe the distributions of the log odds ratio as the measures of the relative treatment effects and then to transform the log odds ratios into more interpretable metrics such as odds ratios or risk ratios. The goal of the comparative effects model is to provide a mechanism to estimate the comparative treatment effects. A critical aspect of the comparative effects model and its relation to network meta-analysis is the consistency assumption. The comparative effects model provides estimates of basic parameters in the form of log odds ratios based on comparisons between each treatment of interest and a baseline treatment. The consistency assumption allows pairwise comparisons between the treatments of interest to be estimated as functions of the basic parameters estimated in the comparative effects model. This consistency assumption is written as:

$$d_{k_1,k_2} = d_{bk_2} - d_{bk_1},$$

where b is the baseline treatment, k_1 and k_2 are treatments other than the baseline, and d_{bk_2} is the true effect size (log odds ratio in this case) of treatment k_2 compared with the baseline b. In lay terms, using the example of bovine respiratory disease, the consistency assumption says that we can compare the effect of oxytetracycline (k_2) with that of tulathrymycin (k_1) if we have comparisons of the effects of oxytetracycline (k_2) and a placebo (b) and of tulathrymycin (k_1) and a placebo (b).

3.2.1. The Fixed Effects Model and the Random Effects Model

The first factor to consider in the comparative effects model is whether the intervention effects are fixed effects or random effects. Suppose there are N studies in a network, which is composed of K treatments. Let b denote the baseline treatment of the whole network, and let b_i denote the trial-specific baseline treatment in trial i. It might be the case that $b_i \neq b$. In other words, the baseline treatment of the model is a placebo, because most of the studies include a placebo group, but a few studies lack a placebo arm and therefore use a different treatment as the baseline comparator. Let y_{ib_ik} be the trial-specific log odds ratio of treatment k compared with b_i in trial i, and let V_{ib_ik} be its within-trial variance. Assume a normal distribution for y_{ib_ik}, such that

$$y_{ib_ik} \sim \mathrm{N}(\theta_{ib_ik}, V_{ib_ik}).$$

The difference between a fixed effects model and a random effects model lies in the assumptions about the nature of the between-trial variability (24). The choice of the fixed effects or random effects model depends on the interpretation of the log odds ratio (θ_{ib_ik}) and the assumptions behind that interpretation. A fixed effects model assumes that there is one true effect size underlying the trials for each comparison. It follows that all of the differences in the observed effect sizes are due to random variation (sampling error) (25), which is akin to assuming that if all the studies were of infinite size, each would result in the same effect size. In that scenario, under the consistency assumption, the model would be:

$$\theta_{i,b_ik} = d_{b_ik} = \begin{cases} d_{bk}, & \text{for } b_i = b, \\ d_{bk} - d_{bb_i}, & \text{for } b_i \neq b, \end{cases}.$$

In this model, d_{bk} ($k \in \{1, 2, \ldots, K\}$) are called basic parameters, whereas d_{b_ik} ($k \in \{1, 2, \ldots, K\}, b_i \neq b$) are called functional parameters, because they are a function of the basic parameters (e.g., $d_{b_ik} = d_{bk} - d_{bb_i}$). For example, consider a trial ($i = 1$) that compared treatment A with treatment B. We might designate treatment A as the baseline treatment (b) and treatment B as k. The model assumes that the log odds ratio observed in study $i = 1$ is d_{bk}. Any difference between the observed log odds ratio and d_{bk} is assumed to be due to sampling error. In another trial ($i = 2$) that compared treatment B to treatment C, we might designate treatment C as the baseline treatment (b_i). When modeling the data, we would retain treatment B as k. The model then assumes that the observed log odds ratio in study $i = 2$ (i.e., treatment C compared with treatment B) is given by d_{bk} - d_{bk_i}. Again, any difference between the observed log odds ratio and d_{bk_i} is assumed to be due to sampling error in a fixed effects model.

A random effects model, on the other hand, assumes that the true effect size can differ from trial to trial, because the effect sizes in each trial are derived from a distribution of effect sizes, which is akin to saying that even if the studies were all of infinite size, there would still be different estimates of the

effect size due to the distribution of effect sizes in addition to sampling error. Therefore, in a random effects model, there is an additional source of variation that needs to be accounted for, that is, the between-trial variation. The random effects model has been recommended for cases in which there is heterogeneity among the results of multiple trials (26). The common distribution of the between-trial variation is usually assumed to be a normal distribution (7), so that

$$\theta_{i,b_ik} \sim \begin{cases} \mathrm{N}\left(d_{bk}, \sigma^2_{b_ik}\right), & \text{for } b_i = b, \\ \mathrm{N}\left(d_{bk} - d_{bb_i}, \sigma^2_{b_ik}\right), & \text{for } b_i \neq b, \end{cases},$$

where $\sigma^2_{b_ik}$ is the between-trial variance. In a pairwise meta-analysis, because there is only one effect size of interest, there is inherently only one between-trial variance. By contrast, in a network meta-analysis, there are at least two, and often many more, effect sizes, because we have ($k \in \{1, 2, \ldots, K\}$). It is often assumed, however, that there is still only a single between-trial variance for all the treatments, which is referred to as the homogeneous variance assumption (i.e., $\sigma^2_{b_ik} = \sigma^2$). In lay terms, this means that if we employ a random effects model that has three treatments and therefore two effect sizes, we assume the same $\sigma^2_{b_ik}$ for d_{bk_1} and d_{bk_2}. Although models that allow heterogeneous between-trial variances have been proposed (4, 27), we use a random effects model with an assumption of homogeneous variance as our example in this tutorial, because such a model is consistent with our biological understanding of the types of interventions used in veterinary science.

3.3. Handling Multi-Arm Trials

In a pairwise meta-analysis, only one effect size is obtained from each study, which means that each effect size is independent of the others. However, in a network meta-analysis, there is the potential, and often the desire, to include multi-arm trials, which creates non-independent observations. For example, a single trial might compare treatments A, B, and C, resulting in three comparisons (A to B, B to C, and B to C). If A is the baseline treatment, then the comparisons between A and B and between A and C are basic parameters. When data from such a trial are included in a network meta-analysis, the assumption of independence is not valid and needs to be adjusted. A term to adjust for the co-variance of data from multi-arm trials must be incorporated into the comparative effects model to correctly reflect the data-generating mechanism. For a single multi-arm trial with k_i treatments, there are $(k_i - 1)$ comparisons $(y_{i,b2}, y_{i,b3}, \ldots, y_{i,bk_i})^T$. The joint distribution of the comparisons is given by

$$\begin{pmatrix} y_{i,b2} \\ y_{i,b3} \\ \vdots \\ y_{i,bk_i} \end{pmatrix} \sim N_{k_i-1}\left(\begin{pmatrix} \theta_{i,b2} \\ \theta_{i,b3} \\ \vdots \\ \theta_{i,bk_i} \end{pmatrix}, \begin{bmatrix} V_{i,b2} & V_{i,b} & \cdots & V_{i,b} \\ V_{i,b} & V_{i,b3} & \cdots & V_{i,b} \\ \vdots & \vdots & \ddots & \vdots \\ V_{i,b} & V_{i,b} & \cdots & V_{i,bk_i} \end{bmatrix}\right),$$

where $V_{i,b}$ is the observed variance in the baseline arm in trial i. The derivation of the value of the co-variance can be found elsewhere (7). For a random effects model, assuming a homogeneous between-trial variance for all trial-specific effects, the joint distribution of $(\theta_{i,b2}, \theta_{i,b3}, \ldots, \theta_{i,bk_i})^T$ is

$$\begin{pmatrix} \theta_{i,b2} \\ \vdots \\ \theta_{i,bk_i} \end{pmatrix} \sim N_{k_i-1}\left(\begin{pmatrix} d_{b2} \\ \vdots \\ d_{bk_i} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 & \ldots & \sigma^2/2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2/2 & \sigma^2/2 & \cdots & \sigma^2 \end{pmatrix}\right).$$

The reason that the off-diagonal values in the variance–covariance matrix are equal to half the diagonal values (28) (i.e., the correlation is 0.5) is that we want to keep the assumption of homogeneous between-trial variance valid. For example,

$$\begin{aligned} \mathrm{Var}(\theta_{i,23}) &= \mathrm{Var}(\theta_{i,b3} - \theta_{i,b2}) = \mathrm{Var}(\theta_{i,b3}) + \mathrm{Var}(\theta_{i,b2}) - 2\mathrm{Cov}(\theta_{i,b3}, \theta_{i,b2}) \\ &= \sigma^2 + \sigma^2 - 2 * \sigma^2/2 = \sigma^2. \end{aligned}$$

3.4. Choice of Priors

So far, we have described the comparative effects model, which describes how the data were generated. The next step is to estimate the parameters of the distributions of interest, that is, the basic parameters for each treatment and the between-trial variance. For a frequentist approach, model parameters are regarded as unknown fixed population characteristics (14) and estimation could be performed using a likelihood approach. The frequentist approach does not use prior information to estimate the parameters. By contrast, the Bayesian approach to estimation calculates the posterior distribution of the parameters by using the data (likelihood) to update prior information. In the Bayesian approach, it is necessary obtain a prior distribution of the parameters, so that the prior distribution can be updated to give the posterior distribution.

Prior distributions must be selected for the basic parameters d_{bk} ($k \in \{1, 2, \ldots, K\}$) and, if a random effects model is employed, also the between-trial variance σ^2. There is no need to select a prior for the correlation of multi-arm trials, because that correlation is constrained to 0.5 by the homogeneous variance assumption. Vague or flat priors such as $\mathrm{N}(0, 10,000)$ are recommended for the basic parameters (7). However, The induced prior on odds ratio (OR), has a big probability on an unrealistic region of odds ratio such that $\Pr(\mathrm{OR} > 1{,}000) \approx 0.47$ and $\Pr(\mathrm{OR} > 10^{29}) \approx 0.25$. However, it provides vague information on the realistic region of the odds ratio and as a result, the posterior distribution depends little on such prior distribution (29). There is no strict rule for selecting a prior for σ^2. The general practice is to set weakly informative priors, such as $\sigma \sim \mathrm{Unif}(0, 2)$ or $\sigma \sim \mathrm{Unif}(0, 5)$, or non-informative priors such as $1/\sigma^2 \sim \mathrm{Gamma}(0.001, 0.001)$. In cases where the data are insufficient, a non-informative prior for σ^2 would be likely to make the posterior distribution include extremely large or small values (30, 31). Lambert et al. (31) conducted a simulation study using 13 vague priors and found that the use of different vague prior distributions led to markedly different results, particularly in small studies. On the basis of those results, Lambert et al. (31) suggested that in any Bayesian analysis, researchers should assess the sensitivity of the results to the choice of the prior

```
model{
        for(i in 1:NS){

           w[i,1] <- 0
           delta[i,t[i,1]] <- 0
           mu[i] ~ dnorm(0, hyperVarInNormMean)

           for(k in 1:na[i]){
             r[i,k] ~ dbin(p[i,t[i,k]], n[i,k])
             logit(p[i,t[i,k]]) <- mu[i] + delta[i,t[i,k]]
             rhat[i,k] <- p[i,t[i,k]] * n[i,k] # expected value of the numerators
             dev[i,k] <- 2*( r[i,k] * (log(r[i,k]) - log(rhat[i,k])) +
                         (n[i,k] - r[i,k]) * (log(n[i,k] - r[i,k]) - log(n[i,k] - rhat[i,k]))
                            )
           }

           resdev[i] <- sum(dev[i,1:na[i]])

           for(k in 2:na[i]){
             delta[i,t[i,k]] ~ dnorm(md[i, t[i,k]], taud[i, t[i,k]])
             md[i, t[i,k]] <- d[t[i,k]] - d[t[i,1]] + sw[i,k]
             taud[i, t[i,k]] <- tau * 2 * (k-1)/k
             w[i,k] <- (delta[i,t[i,k]] - d[t[i,k]] + d[t[i,1]])
             sw[i,k] <- sum(w[i, 1:(k-1)])/(k-1)
           }
        }

        totresdev <- sum(resdev[])

        d[1] <- 0
        for(k in 2:NT){
            d[k] ~ dnorm(0, hyperVarInNormMean)
        }

        sd ~ dunif(0,hyperSDInUnif)
        tau <- 1/pow(sd,2)
```

FIGURE 1 | An example of the formatting of the BUGS code for the comparative model. This code was modified from code originally published elsewhere (7).

distribution for σ^2, because "vague" is not the same in all cases. For example, if the prior chosen for the between-study variance is Unif(0,5), a sensitivity analysis for that prior could look at how the posterior estimates of the treatment effects (e.g., the log odds ratios or the absolute risks) in the network meta-analysis would change if the prior is changed to Unif(0,2), Unif(0,10), or some other distribution. If the posterior estimates do not change substantially, the results can be considered insensitive to the choice of prior parameter values. Informative priors can be considered if there are reasonable estimates of σ^2 available from another, larger network meta-analysis that has the same context and similar treatments as the analysis under construction (7, 32). Having considered the choice of a random effects or fixed effects model, the handling of multi-arm trials, and the choice of priors, the specification of the comparative effects model is complete. **Figure 1** illustrates an example of the coding of a comparative effects model in the general_model.bug code.

3.5. The Baseline Effects Model–the Log Odds of the Event

After defining the comparative effects model and the priors for the parameters to be estimated, the next step in a Bayesian network meta-analysis is to model the baseline effect. Although it is possible to conduct a Bayesian network meta-analysis without a baseline effects model, the baseline effects model allows for some unique and informative outputs from the analysis. If we are only interested in the estimates of the log odds ratios and the odds ratios, then there is no need to make a baseline effects model. The baseline effects model refers to the distribution of the event for the baseline treatment, that is, the log odds of the event for the baseline treatment. For example, if in one study 40 out of 100 animals in the baseline group experienced the event, the trial-specific log odds of the event would be $\log((\frac{40}{100})/(\frac{60}{100}))$. A different trial would have different log odds of the event; however, the log odds of the event are assumed to arise from the same distribution in all trials. The reason for modeling the distribution of the event risk in the baseline group is to enable absolute effects (i.e., absolute risk) and comparative effects to be estimated on a risk scale rather than on an odds scale. For example, if we know the log odds ratio for all treatments compared with the baseline treatment, then, given the absolute risk for any one treatment, we can know the absolute risk for every treatment. For example, if we have a log odds ratio of 0.9809 for the comparison between treatment A and the baseline treatment, and if the baseline event risk is 0.2 (e.g., 20 out of 100 exposed subjects experienced the event), then we can determine that the absolute risk for treatment

A is 0.4, using the formula

$$p = \frac{\text{OR} \times p_b}{1 - p_b + \text{OR} \times p_b},$$

where p_b is the absolute risk for the baseline treatment, and p is the absolute risk for any non-baseline treatment. The absolute effect of the baseline treatment is often selected for baseline effect modeling, because the baseline treatment is usually the most common treatment in the network meta-analysis, which means that it has the most data available for estimation of the posterior distribution of the log odds of the event. Suppose there are N_b studies that have the baseline arm. Let $\theta_{i,b}$ ($i \in \{1, \ldots, N_b\}$) be the trial-specific baseline effect (log odds of the event) in a trial i (i.e., the log odds). We can use the following formulation to model the baseline effect:

$$\theta_{i,b} \sim \text{N}(m, \sigma_m^2).$$

This means that the trial-specific baseline effects come from a normal distribution with mean m and variance σ_m^2. As with the comparative effects model, we need to select priors for the baseline effects model. The selection of prior distributions for m and σ_m^2 follows the same considerations as the selection of priors for the effect parameters, that is, the priors should be weakly informative or non-informative [e.g., $m \sim \text{N}(0, 10000)$, and $\sigma_m \sim \text{Unif}(0, 5)$]. From a coding perspective, there are two ways to incorporate a baseline effects model into the comparative effects model. The first approach is to run separate models, beginning with the baseline effects model. The baseline model yields the posterior distribution summaries of m and σ_m (or σ_m^2). The posterior means (denoted by $\hat{m}, \hat{\sigma}_m^2$) are then inserted into the comparative effects model and the baseline effect can be generated from $\text{N}(\hat{m}, \hat{\sigma}_m^2)$ in the comparative effects model. Other quantities of interest (e.g., the absolute risk for the other treatments) can then be estimated. The first approach relies on the assumption that the posterior distribution of the baseline effect is approximately normal. Dias et al. (33) suggests checking that assumption (e.g., with Q-Q plot or Kolmogorov–Smirnov test), although the assumption is usually found to hold. The second approach to incorporate the baseline effects model into the comparative effects model is simultaneous modeling of the baseline effect and the comparative effects. That approach can have a substantial impact on the relative effect estimates, however. For more details on the simultaneous modeling of baseline and comparative effects, refer to Dias et al. (33). **Figure 2** shows the incorporation of a baseline effects model, which can be used to obtain m and σ_m (or σ_m^2).

```
model{

for (i in 1:ns){
    r[i] ~ dbin(p[i],n[i])
    logit(p[i]) <- mu[i]
    mu[i] ~ dnorm(m,tau.m)
  }


mu.new ~ dnorm(m,tau.m)

m ~ dnorm(0, hyperVarInNormMean)
sd.m ~ dunif(0, hyperSDInUnif)
tau.m <- pow(sd.m,-2)
```

FIGURE 2 | An example of the formatting of the BUGS code for the baseline effects model. This code was modified from code originally published elsewhere (7).

4. THE WORKFLOW FOR CONDUCTING A BAYESIAN NETWORK META-ANALYSIS

4.1. Data Input

The data used in network meta-analyses are typically arranged in one of three formats: one study per row, one comparison per row (contrast-level data), or one arm per row (arm-level data only). In our network meta-analysis functions, we use the one-study-per-row format. The example data that we use in the following analysis are shown in arm-level format in **Table 1**. In the example data, there are five treatments (A, B, C, D, and E). The baseline treatment is A. It is essential that the baseline treatment is indexed as one (1) and that the data are organized such that the baseline treatment arms are always the "Arm1" treatment. If there are trials with more than two arms, then corresponding columns (e.g., "Number of Events in Arm.1," "Arm.3," "Arm3") can simply be added to the dataset. **Table 2** shows the same data arranged in contrast-level format.

4.2. Running the Analysis

After we select studies that meet the transitivity assumption, extract the data and arrange them in the necessary format, decide upon a fixed or random effects model, set the priors for the basic parameters, determine the boundaries of the between-trial variance based on the data, and obtain $\hat{m}$ and $\hat{\sigma}_m$ (or $\hat{\sigma}_m^2$) from the baseline effects model, the next step is to run the network meta-analysis.

4.3. A Description of the Workflow of a Network Meta-Analysis

The workflow of a Bayesian network meta-analysis can be described as follows:

1. Use the comparative effects model and a Markov chain Monte Carlo (MCMC) process to obtain the posterior distributions of the log odds ratios for the basic parameters. From those basic parameters, obtain the posterior distributions of the functional parameters. After running the model the next sub-steps are to:
    a. Assess the convergence by evaluating the trace plots and convergence criteria such as the potential scale reduction factor proposed by Gelman and Rubin (34).

TABLE 1 | Example data arranged in arm-level format.

Study	Number of event in arm.1	Number of event in arm.2	Number of event in arm.3	Total number in arm.1	Total number in arm.2	Total number in arm.3	Total	Arm.1	Arm.2	Arm.3	Number of arms	Arm1	Arm2	Arm3
1	25	17	20	41	84	100	225	A	B	C	3	1	2	3
2	36	32		41	84		125	A	B		2	1	2	
3	19	7		25	25		50	A	B		2	1	2	
4	20	5		25	50		75	A	B		2	1	2	
5	41	47		50	100		150	A	B		2	1	2	
6	122	69		160	314		474	A	E		2	1	5	
7	236	53		402	399		801	A	E		2	1	5	
8	23	15		27	52		79	A	E		2	1	5	
9	175	166		281	274		555	B	E		2	2	5	
10	57	20		119	118		237	B	E		2	2	5	
11	19	12		100	100		200	B	E		2	2	5	
12	19	7		100	100		200	B	E		2	2	5	
13	16	21		258	254		512	B	E		2	2	5	
14	42	15		50	100		150	A	B		2	1	2	
15	64	34		154	154		308	A	C		2	1	3	
16	34	15		53	106		159	A	C		2	1	3	
17	70	42		130	129		259	A	C		2	1	3	
18	92	31		121	121		242	A	C		2	1	3	
19	35	20		45	90		135	A	C		2	1	3	
20	41	62		59	117		176	A	C		2	1	3	
21	37	15		43	85		128	A	C		2	1	3	
22	16	21		18	35		53	A	C		2	1	3	
23	70	35		122	123		245	A	B		2	1	2	
24	204	71		300	300		600	A	D		2	1	4	
25	111	66		523	526		1049	C	E		2	3	5	
26	60	50		305	297		602	B	C		2	2	3	

The last two columns are the treatment indexes used to distinguish different treatments in the code.

b. Check the goodness of the model's fit using the (residual) deviance. It is the posterior mean of the difference in the negative 2× log likelihood between the current model and the saturated model (35). An empirical rule to check if the model fits well (7) is that the value of the residual deviance should be close to the number of independent data points (36).
c. Obtain the summary information [mean, standard deviation (*SD*)] of the distributions of basic parameters and functional parameters from the comparative effects model and also the summary information (mean, *SD*) of the distributions of basic parameters from the pairwise comparative effects model.

2. Use pairwise comparative effects models and the MCMC process to obtain the posterior distribution of the log odds ratio for the treatments that have direct comparisons that can be used later to check the consistency assumption. This step is essentially a series of Bayesian pairwise meta-analyses based on direct estimates. Hence, no indirect evidence is used in the estimation procedure. After running the model again the next sub-steps are to:
   a. Ensure convergence by evaluating the trace plots and convergence criteria.
   b. Obtain the summary information of the distributions of basic parameters and functional parameters from the pairwise comparative effects model.
3. Using data from Step 1 and 2, assess the consistency assumption for the treatment comparisons for which there is direct evidence. This is done by subtracting the mean estimated log odds ratios obtained from the posterior distributions of the pairwise meta-analyses from the mean estimated log odds ratios obtained from the posterior distributions of the network meta-analysis and looking for inconsistencies (37). The "indirect estimates" can be obtained by

$$\hat{d}_{\text{indir}} = \text{Var}(\hat{d}_{\text{indir}})\left(\frac{\hat{d}_{\text{NMA}}}{\text{Var}(\hat{d}_{\text{NMA}})} - \frac{\hat{d}_{\text{dir}}}{\text{Var}(\hat{d}_{\text{dir}})}\right),$$

$$\frac{1}{\text{Var}(\hat{d}_{\text{indir}})} = \frac{1}{\text{Var}(\hat{d}_{\text{NMA}})} - \frac{1}{\text{Var}(\hat{d}_{\text{dir}})},$$

and should be consistent with the direct estimates. For example, if the pairwise comparison of treatment A with treatment B gives a mean difference in effect size of 1.2, then the indirect comparison of those treatments should give a mean difference in effect size that is positive and of similar magnitude. The hypothesis that the difference between the direct and indirect estimates is zero can be tested using a z-score and corresponding p-value. Such hypothesis tests are often very low powered, however, so it is recommended to also visually evaluate the magnitude and direction of the indirect effects and determine if they are consistent with the direct effects.

If there is no evidence of inconsistency, and residual deviance is also not a concern, then the network meta-analysis is complete. If there is inconsistency, then it is necessary to evaluate the included studies to determine the cause of the inconsistency. In our experience, we once identified an issue with inconsistency that appeared to be linked to a single study that contained results that were not consistent with those of the other studies in the network. In that situation, we removed the problematic study from the network and performed the network meta-analysis without it. More information about that example can be found elsewhere (6).

The next step is to convert the distributional information about the basic and functional parameters into a form that is appropriate for presentation and interpretation. First, we will discuss the estimates of the treatment effects (i.e., the log odds ratios, odds ratios, and risk ratios). Then, we will discuss how to derive information from those estimates. In

TABLE 3 | The estimated log odds ratio from all possible pairwise comparisons in the network meta-analysis of five treatment groups.

E	−0.648	−0.689	−0.475	−2.576
(−2.304_0.983)	**D**	−0.041	0.174	−1.928
(−1.394_0.017)	(−1.646_1.559)	**C**	0.214	−1.887
(−1.058_0.108)	(−1.421_1.797)	(−0.422_0.850)	**B**	−2.101
(−3.208_-1.969)	(−3.451_-0.415)	(−2.404_−1.398)	(−2.653_−1.577)	**A**

All the point estimates are the posterior mean of the log odds ratio of the upper left treatment to the lower right treatment. For example, −2.101 is the posterior mean of the log odds ratio of treatment B to treatment A. (−2.653_−1.577) is the 95% credible interval of the log odds ratio of treatment B to treatment A.

TABLE 2 | Example data in contrast-level format.

Study	Arm.1	Arm.2	Arm.3	Number of arms	lor 2	lor 3	se 2	se 3	Arm1	Arm2	Arm3	V	PLA lo
1	A	B	C	3	−1.82	−1.83	0.42	0.41	1	2	3	0.10	0.45
2	A	B		2	−2.46		0.53		1	2			1.97
3	A	B		2	−2.10		0.65		1	2			1.15
4	A	B		2	−3.58		0.69		1	2			1.39
5	A	B		2	−1.64		0.42		1	2			1.52
6	A	E		2	−2.43		0.23		1	5			1.17
7	A	E		2	−2.23		0.18		1	5			0.35
8	A	E		2	−2.65		0.62		1	5			1.75
9	B	E		2	−0.07		0.17		2	5			
10	B	E		2	−1.51		0.31		2	5			
11	B	E		2	−0.54		0.40		2	5			
12	B	E		2	−1.14		0.47		2	5			
13	B	E		2	0.31		0.34		2	5			
14	A	B		2	−3.39		0.48		1	2			1.66
15	A	C		2	−0.92		0.25		1	3			−0.34
16	A	C		2	−2.38		0.40		1	3			0.58
17	A	C		2	−0.88		0.26		1	3			0.15
18	A	C		2	−2.22		0.30		1	3			1.15
19	A	C		2	−2.51		0.44		1	3			1.25
20	A	C		2	−0.70		0.34		1	3			0.82
21	A	C		2	−3.36		0.52		1	3			1.82
22	A	C		2	−1.67		0.83		1	3			2.08
23	A	B		2	−1.22		0.27		1	2			0.30
24	A	D		2	−1.92		0.18		1	4			0.75
25	C	E		2	−0.63		0.17		3	5			
26	B	C		2	−0.19		0.21		2	3			

"lor 2" is the column of log odds ratio of "Arm 2" to "Arm 1." "se 2" shows the corresponding within-trial standard error. The column labeled "V" contains the variance of the log odds of "Arm 1" only if the trial has more than two arms, as discussed in the section "Multi-arm trials." The column labeled "PLA lo" contains the log odds for the baseline treatment.

reality, the distributions of the treatment effects are obtained during the performance of the network meta-analysis. When the MCMC process is conducted, each simulation yields an odds ratio, a baseline event risk, and a risk ratio. The posterior distributions of the parameters and the summary statistics for the distributions are then extracted from the raw data produced by the simulations. Thus, it is possible to report the following:

- All possible log odds ratios with 95% credible intervals as shown in **Table 3**. These are estimated from the model using the indirect and direct information.
- All possible pairwise odds ratios with 95% credible intervals (**Table 4**). These are estimated by converting each log odds ratio to an odds ratio during each simulation and then obtaining the posterior distribution of the odds ratios. These cannot be obtained by exponentiation of the mean or the limits of the posterior distribution of the log odds ratio.
- All possible pairwise risk ratios with 95% credible intervals (**Table 5**). These estimates are obtained for each simulation by using the basic parameters (log odds ratios) and the baseline risk to calculate the probability of an event for each treatment with the expit formula. For example, if for a particular simulation the log odds ratio for treatment B compared with treatment A is 0.9809 (odds ratio of 2.667), and the baseline risk for treatment A is 20%, then the risk of an event for treatment B is 40%. The treatment event risks are then used to create risk ratio estimates (40/20%).

Apart from estimating all possible pairwise treatment effects using direct and indirect data on different scales, it is also possible to create other outputs that help to illustrate aspects of the data.

TABLE 4 | The estimated odds ratio from all possible pairwise comparisons in the network meta-analysis of five treatment groups.

E	0.743	0.535	0.650	0.080
(0.100_2.672)	**D**	1.347	1.678	0.196
(0.248_1.017)	(0.193_4.753)	**C**	1.305	0.157
(0.347_1.114)	(0.241_6.033)	(0.656_2.341)	**B**	0.127
(0.040_0.140)	(0.032_0.660)	(0.090_0.247)	(0.070_0.207)	**A**

All the point estimates are the posterior mean of the log odds ratio of the upper left treatment to the lower right treatment. For example, 0.127 is the posterior mean of the odds ratio of treatment B to treatment A.

TABLE 5 | The estimated risk ratio from all possible pairwise comparisons in the network meta-analysis of five treatment groups with the summary of baseline risk to be mean = 0.713, median = 0.728, 2.5% limit = 0.45, 97.5% limit = 0.899.

E	0.781	0.616	0.711	0.252
(0.208_2.309)	**D**	1.074	1.260	0.423
(0.326_1.012)	(0.263_2.543)	**C**	1.200	0.411
(0.422_1.083)	(0.310_3.059)	(0.736_1.894)	**B**	0.356
(0.102_0.496)	(0.094_0.900)	(0.205_0.675)	(0.168_0.621)	**A**

All the point estimates are the posterior mean of the risk ratio of the upper left treatment to the lower right treatment. For example, 0.356 is the posterior mean of the risk ratio of treatment B to treatment A.

TABLE 6 | Summary of the distribution of the rankings for the five treatments.

Treatment	Mean	*SD*	2.5%	50%	97.5%
A	4.99	0.09	5	5	5
C	3.25	0.74	2	3	4
D	2.86	1.21	1	3	4
B	2.60	0.75	1	3	4
E	1.29	0.54	1	1	3

TABLE 7 | The probability of being the best treatment and the probability of being the worst treatment.

Treatment	Probability of being best	Probability of being worst
A	0.000	0.992
B	0.033	0.000
C	0.015	0.000
D	0.201	0.008
E	0.751	0.000

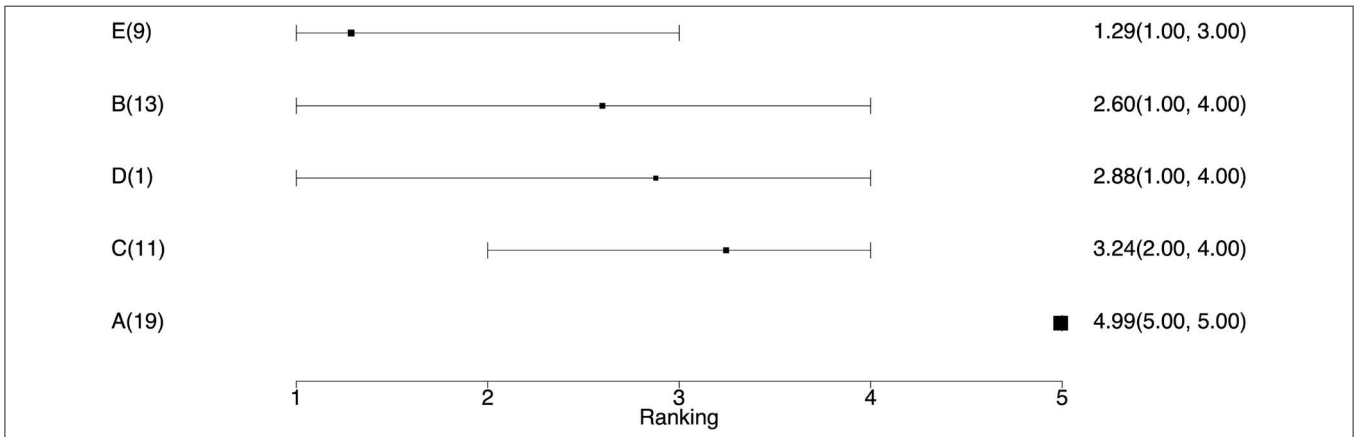


FIGURE 3 | The ranking plot. The left column is the treatment name with the number of studies including that treatment. The right column is the posterior mean ranking of the absolute risk of each treatment and 95% credible interval. Lower rankings have lower incidence of the disease.

There are many options, but here we discuss only a few. Many outputs are based on the creation of an indicator variable that takes a given value at a frequency proportional to the probability of an event. The indicators can be created during the simulation process or *post-hoc* in R. The code in the Appendix provides examples of both approaches.

- The average ranking of each treatment (**Table 6**). Once the event probability has been determined for each simulation, it is then possible to rank the event risk across all the treatments. A numerical value ranging from 1 to the total number of treatments is then assigned to each treatment. The researcher can determine what is considered a good or high rank based on the event and what value to assign the most desirable rank. Usually, a rank of 1 is assigned as the preferred result. For example, consider one simulation where the probability of an event for treatments A, B, C, D, and E is 10, 15, 17, 20, and 30%, respectively. If the event is a desirable characteristic, such as a cure, then the treatments A, B, C, D, and E would be assigned the ranks 5, 4, 3, 2, and 1, respectively. In the next simulation, the probability of the event for treatments A, B, C, D, and E might be 5, 22, 17, 24, and 33%, respectively, so treatments A, B, C, D, and E would be ranked as 5, 3, 4, 2, and 1, respectively. In a Bayesian analysis, the posterior samples from all three chains can be used to create a posterior distribution of the rankings. The summary statistics of the posterior distribution of the rankings can be reported. Often the mean or median of the posterior distribution of the rankings and the 95% credible intervals of the rankings are used to create a ranking plot, as shown in **Figure 3**.
- The probability of being the best (or worst) treatment (**Table 7**). Using the data from the rankings, it is possible to sum the number of times each treatment received the highest (or lowest) rank. The sum can then be reported as the probability that the treatment has the highest (or lowest) rank, which is colloquially interpreted as the probability of being the best (or worst) treatment.
- All possible pairwise comparisons of the probability of being better (**Table 8**). Using the ranking data, which are based on the event risk data for each treatment, it is possible to

TABLE 8 | The probability that one treatment is better than another, i.e., has lower disease incidence during the study period.

A	0.000	0.000	0.008	0.000
1.000	**B**	0.757	0.587	0.052
1.000	0.243	**C**	0.476	0.028
0.992	0.413	0.524	**D**	0.206
1.000	0.948	0.972	0.794	**E**

The upper quadrant provides the probability that the row treatment is better than the column. For example, there is probability of zero that "A" (1st row) is better than "B" (2nd column) and a probability of 0 that "A" is better than "E".

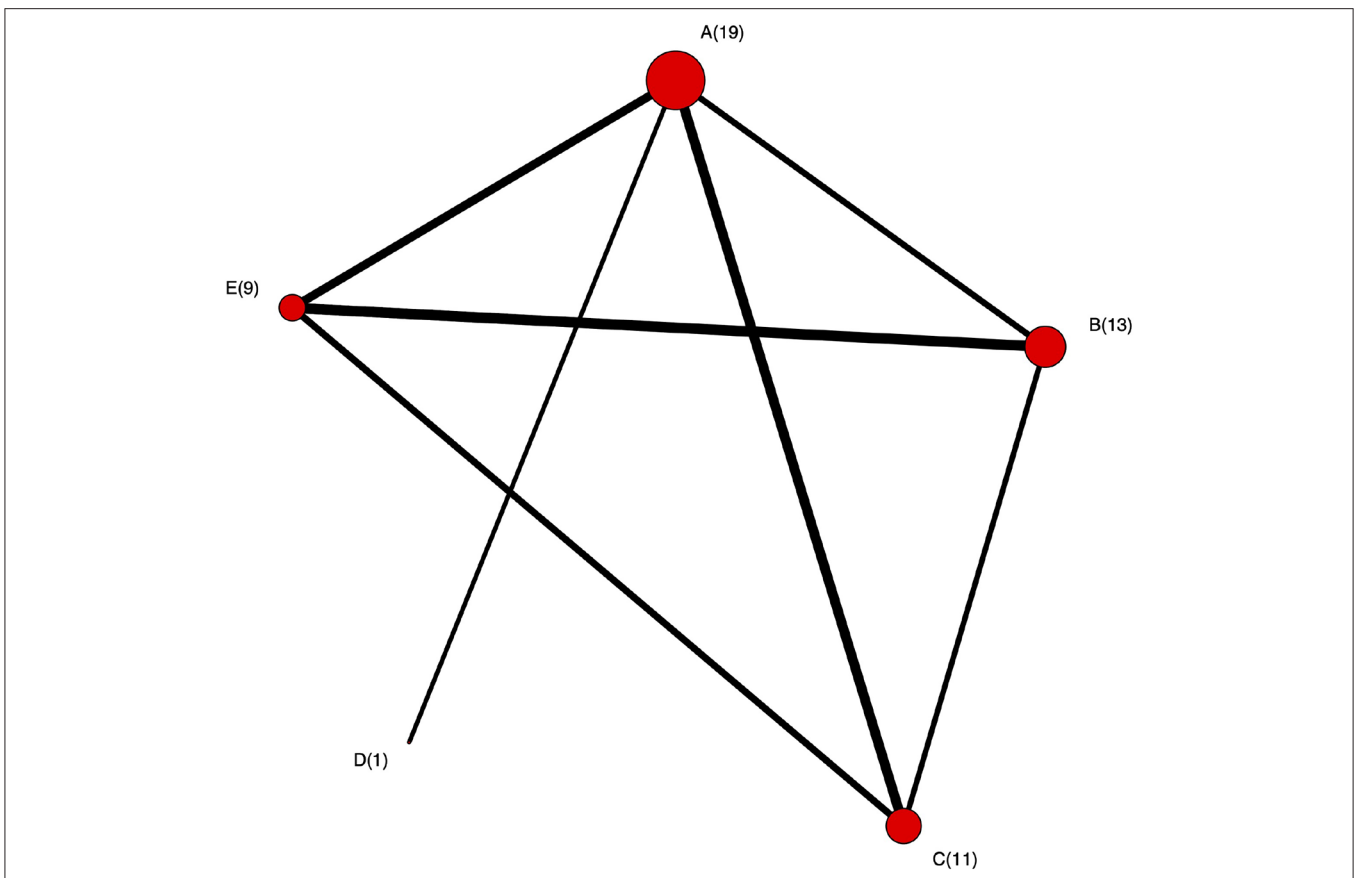


FIGURE 4 | The network plot. Each node represents treatment and the number is the corresponding number of studies including that treatment. An edge between two nodes (treatments) means there were studies comparing these two treatments.

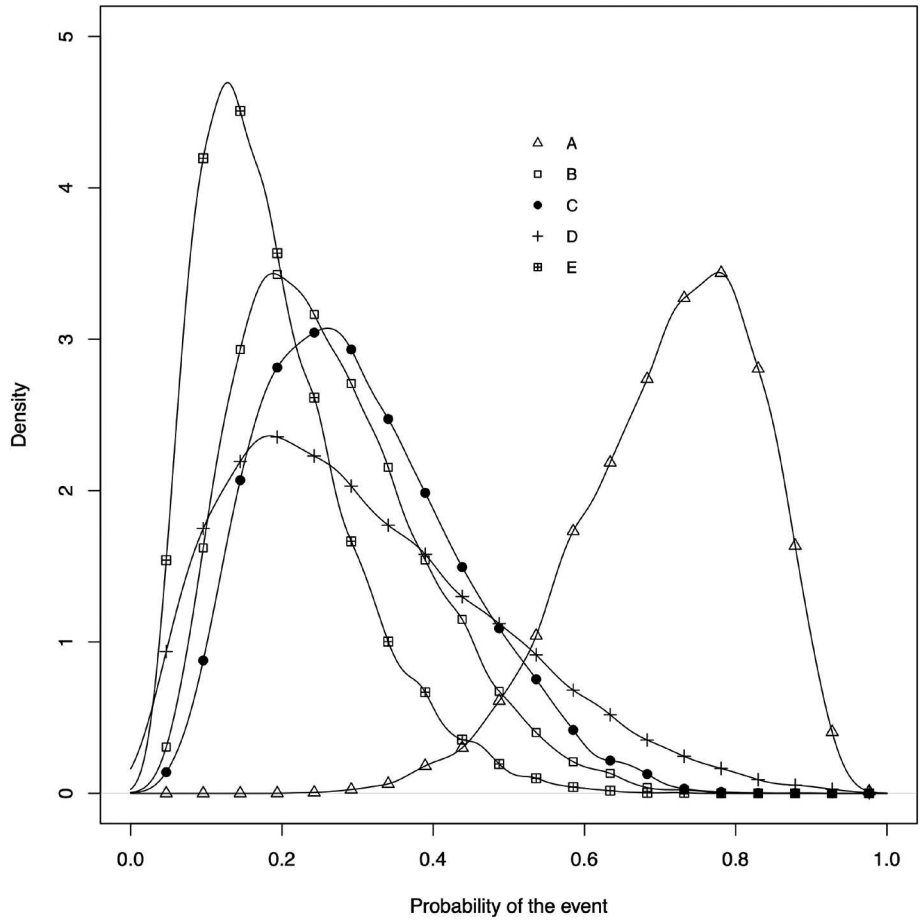


FIGURE 5 | The posterior distribution of the event risk of each treatment.

sum the proportion of times that one treatment is ranked higher (or has a higher event rate) than another treatment. This can be done using either the ranking data or the event risk data, which both give the same result. In our example data, the probability that B, C, D, and E were better than A was 10%, whereas the probability that B was better than C was 50%.

4.4. Plots Commonly Used to Show the Results of a Network Meta-Analysis

There are various types of plots that can be used to present the results of a network meta-analysis. Examples of three of the most common types are shown below.

- The network plot as shown in **Figure 4**. This plot is a visual representation of the network of evidence. Although we did not discuss the network plot until the end of the tutorial, because it is not technically part of the network meta-analysis, this plot should actually be generated before the network meta-analysis is undertaken. The code provided in the Appendix illustrates how to create the network plot using packages from R. There are also other approaches that can be used to create the network plot. The code in the Appendix includes some common metrics used to describe networks, which are not discussed further here (38).
- The posterior distribution of the event risk (**Figure 5**). This plot illustrates the posterior distribution of the event risk for each treatment using all posterior samples of that risk.
- The ranking plot (**Figure 3**). The ranking plot uses the data from the posterior distribution of the rankings to create a forest plot-like graphic using the means and 95% credible intervals of the rankings.

5. DISCUSSION

In this tutorial, we described the conceptual framework for a network meta-analysis, explained the step-wise workflow for conducting a network meta-analysis, and provided code in the Appendix that illustrates the mechanics of conducting a Bayesian network meta-analysis. The Bayesian inference tool used in this tutorial is JAGS. Stan Development Team (39), as an alternative Bayesian inference instrument, could also be used to conduct network-meta analysis. As we mentioned in

the introduction section, other packages for network meta-analysis like *gemtc* and *BUGSnet* are also available. Compared with *gemtc*, the outputs in our code are more flexible and are shown in table format. Our code can also deal with arm-level data as well as contrast-level data in comparison to *BUGSnet*. Despite these advantages, there are some limitations. Our code focuses on the binary outcome. *gemtc* and *BUGSnet* provide functions handling other types of outcome like continuous and count outcomes.

Network meta-analysis, as a popular method of simultaneously comparing multiple treatments, still presents challenges since it not only has the challenges as in a standard pairwise meta-analysis but also increases the complexity due to the network structure (40). Therefore, some assumptions are made to ensure the validness of a network meta-analysis. The transitivity assumption is that studies can be combined only when they are clinically and methodologically similar (41, 42). This means according to the Cochrane Handbook of Systematic Reviews "that different sets of randomized trials are similar, on average, in all important factors other than the intervention comparison being made" (43). For example, the distributions of effect modifiers should be similar across studies (44). Practically, in our BRD example, the transitivity assumption means that each study population would have been eligible for any of the other studies and all study populations were eligible for all treatments. An example of a situation that would violate this transitivity assumption would be a comparison of antibiotic treatment efficacy where one group of trials assessed the response to 1st treatment and another group of trials assessed the treatment response of cattle with a 1st treatment failure (re-pull). Obviously, the cattle in the 1st treatment response are not eligible for the 1st treatment failure studies. The validity of indirect and combined estimates of relative effects would be threatened if this assumption is violated (43). Consistency assumption is a manifestation of transitivity. As we discussed in section 3.2, it requires that the indirect evidence must be consistent with direct evidence. Violation of the consistency assumption would result in inconsistency (45). Although inconsistency model have been proposed to mitigate the violation of this assumption in some way, one still should be cautious when combining studies and choosing which model to use. This tutorial focuses on the statistical aspect of conducting a network meta-analysis while aspects such as defining the research question, searching for studies and assessing the risk of bias within each study (46, 47) are not in the scope of this tutorial.

For readers that are interested in running a simple network meta-analysis without going into any detailed explanation of the underlying process, we believe that the instructions that come with any one of the ever-growing number of software packages for network meta-analysis will provide sufficient information for a successful analysis to be conducted (12, 15–17). More details about interpreting the results of a network meta-analysis can be found on this paper (48).

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the GitHub [https://github.com/a-oconnor/NETWORK_MA_FRONTIERS_TUTORIAL].

AUTHOR CONTRIBUTIONS

DH prepared the draft of the manuscript, wrote the code used to conduct the data analysis, and worked with AO'C to ensure that the interpretation was correct. AO'C prepared the draft of the manuscript, coordinated the project team, assisted with the data analysis, and interpreted the procedure and results of the analysis. CW provided guidance on the conduct of the analysis and commented on the draft of the manuscript to ensure that the interpretation of the analysis was clear. JS provided feedback on the draft to ensure that the interpretation of the analysis was clear. CBW provided feedback on the draft to ensure that the interpretation of the analysis was clear.

REFERENCES

1. Coleman CI, Phung OJ, Cappelleri JC, Baker WL, Kluger J, White CM, et al. *Use of Mixed Treatment Comparisons in Systematic Reviews [Internet]*. Agency for Healthcare Research and Quality (2012). Available online at: www.effectivehealthcare.ahrq.gov/reports/final.cfm
2. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med*. (2002) 21:589–624. doi: 10.1002/sim.1040
3. Smith E, Hubbard S, Cooper N, Abrams K. Using network meta-analysis of individual patient data (IPD) & summary aggregate data (SAD) to identify which combinations of interventions work best for which individuals. *Value Health*. (2017) 20:A758. doi: 10.1016/j.jval.2017.08.2138
4. Lu G, Ades A. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. (2004) 23:3105–24. doi: 10.1002/sim.1875
5. O'Connor AM, Coetzee JF, da Silva N, Wang C. A mixed treatment comparison meta-analysis of antibiotic treatments for bovine respiratory disease. *Prev Vet Med*. (2013) 3:77–87. doi: 10.1016/j.prevetmed.2012.11.025
6. O'Connor AM, Yuan C, Cullen JN, Coetzee JF, da Silva N, Wang C. A mixed treatment meta-analysis of antibiotic treatment options for bovine respiratory disease - An update [Journal Article]. *Prev Vet Med*. (2016) 132:130–9. doi: 10.1016/j.prevetmed.2016.07.003
7. Dias S, Welton NJ, Sutton AJ, Ades A. *NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials.* (2011). Available online at: http://research-information.bristol.ac.uk/files/7215331/TSD2_General_meta_analysis.final.08.05.12.pdf
8. Calzetta L, Roncada P, di Cave D, Bonizzi L, Urbani A, Pistocchini E, et al. Pharmacological treatments in asthma-affected horses: a pair-wise and network meta-analysis. *Equine Vet J*. (2017) 49:710–7. doi: 10.1111/evj.12680
9. Abell KM, Theurer ME, Larson RL, White BJ, Apley M. A mixed treatment comparison meta-analysis of metaphylaxis treatments for bovine respiratory disease in beef cattle. *J Anim Sci*. (2017) 95:626–35. doi: 10.2527/jas2016.1062
10. Jacobs C, Beninger C, Hazlewood GS, Orsel K, Barkema HW. Effect of footbath protocols for prevention and treatment of digital dermatitis in dairy cattle: a systematic review and network meta-analysis. *Prev Vet Med*. (2019) 164:56–71. doi: 10.1016/j.prevetmed.2019.01.011
11. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna (2019). Available online at: https://www.R-project.org/

12. van Valkenhoef G, Kuiper J. *gemtc: Network Meta-Analysis Using Bayesian Methods*. R package version 0.8-2 (2016). Available online at: https://CRAN.R-project.org/package=gemtc
13. Béliveau A, Boyne DJ, Slater J, Brenner D, Arora P. BUGSnet: an R package to facilitate the conduct and reporting of Bayesian network Meta-analyses. *BMC Med Res Methodol*. (2019) 19:196. doi: 10.1186/s12874-019-0829-2
14. Shim S, Yoon BH, Shin IS, Bae JM. Network meta-analysis: application and practice using Stata. *Epidemiol Health*. (2017) 39:e2017047. doi: 10.4178/epih.e2017047
15. Neupane B, Richer D, Bonner AJ, Kibret T, Beyene J. Network meta-analysis using r: a review of currently available automated packages. *PLoS ONE*. (2014) 9:e115065. doi: 10.1371/journal.pone.0115065
16. Staff TPO. Correction: network meta-analysis using r: a review of currently available automated packages. *PLoS ONE*. (2015) 10:e123364. doi: 10.1371/journal.pone.0123364
17. Lin L, Zhang J, Hodges JS, Chu H. Performing Arm-Based Network Meta-Analysis in R with the pcnetmeta Package. *J Stat Softw*. (2017) 80:5. doi: 10.18637/jss.v080.i05
18. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. (2010) 36:1–48. doi: 10.18637/jss.v036.i03
19. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Vol. 124. Vienna (2003). p. 1–10.
20. Gilks WR, Thomas A, Spiegelhalter DJ. A language and program for complex Bayesian modelling. *J R Stat Soc*. (1994) 43:169–77. doi: 10.2307/2348941
21. Sargeant JM, O'Connor AM. Scoping reviews, systematic reviews, and meta-analysis: applications in veterinary medicine. *Front Vet Sci*. (2020) 7:11. doi: 10.3389/fvets.2020.00011
22. O'connor A, Sargeant J, Wang C. Conducting systematic reviews of intervention questions iii: synthesizing data from intervention studies using meta-analysis. *Zoonoses Public Health*. (2014) 61:52–63. doi: 10.1111/zph.12123
23. Cohen J. A power primer. *Psychol Bull*. (1992) 112:155. doi: 10.1037/0033-2909.112.1.155
24. Altman DG, Smith GD, Egger M. Systematic reviews in health care: meta analysis in context. *BMJ*. (2001) 323:224–8. doi: 10.1136/bmj.323.7306.224
25. Hedges LV, Vevea JL. Fixed-and random-effects models in meta-analysis. *Psychol Methods*. (1998) 3:486. doi: 10.1037/1082-989X.3.4.486
26. Greco T, Landoni G, Biondi-Zoccai G, D'Ascenzo F, Zangrillo A. A Bayesian network meta-analysis for binary outcome: how to do it. *Stat Meth Med Res*. (2016) 25:1757–73. doi: 10.1177/0962280213500185
27. Turner RM, Domínguez-Islas CP, Jackson D, Rhodes KM, White IR. Incorporating external evidence on between-trial heterogeneity in network meta-analysis. *Stat Med*. (2019) 38:1321–35. doi: 10.1002/sim.8044
28. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med*. (1996) 15:2733–49. doi: 10.1002/(SICI)1097-0258(19961230)15:24<2733::AID-SIM562>3.0.CO;2-0
29. van Valkenhoef G, Lu G, de Brock B, Hillege H, Ades A, Welton NJ. Automating network meta-analysis. *Res Synth Methods*. (2012) 3:285–99. doi: 10.1002/jrsm.1054
30. Thorlund K, Thabane L, Mills EJ. Modelling heterogeneity variances in multiple treatment comparison meta-analysis-are informative priors the better solution? *BMC Med Res Methodol*. (2013) 13:2. doi: 10.1186/1471-2288-13-2
31. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Stat Med*. (2005) 24:2401–28. doi: 10.1002/sim.2112
32. Ren S, Oakley JE, Stevens JW. Incorporating genuine prior information about between-study heterogeneity in random effects pairwise and network meta-analyses. *Med Decis Making*. (2018) 38:531–42. doi: 10.1177/0272989X18759488
33. Dias S, Welton NJ, Sutton AJ, Ades A. *NICE DSU Technical Support Document 5: Evidence Synthesis in the Baseline Natural History Model*. Technical Support Document, National Institute for Health and Clinical Excellence, London. (2011).
34. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. (1992) 7:457–72. doi: 10.1214/ss/1177011136
35. McCulloch CE, Neuhaus JM. Generalized linear mixed models. *Encyclop Biostat*. (2005) 4:2085–89. doi: 10.1002/0470011815.b2a10021
36. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc*. (2002) 64:583–639. doi: 10.1111/1467-9868.00353
37. Dias S, Welton N, Caldwell D, Ades A. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med*. (2010) 29:932–44. doi: 10.1002/sim.3767
38. Salanti G, Kavvoura FK, Ioannidis JP. Exploring the geometry of treatment networks [Journal Article]. *Ann Intern Med*. (2008) 148:544–53. doi: 10.7326/0003-4819-148-7-200804010-00011
39. Stan Development Team. *RStan: the R interface to Stan*. R package version 2.19.1 (2019). Available online at: http://mc-stan.org/
40. Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med*. (2013) 159:130–7. doi: 10.7326/0003-4819-159-2-201307160-00008
41. Caldwell DM, Ades A, Higgins J. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ*. (2005) 331:897–900. doi: 10.1136/bmj.331.7521.897
42. Dias S, Sutton AJ, Ades A, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making*. (2013) 33:607–17. doi: 10.1177/0272989X12458724
43. Chaimani A, Caldwell DM, Li T, Higgins JP. Undertaking network meta-analyses. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.0 (updated July 2019)*. London: Cochrane (2019) p. 285–320.
44. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods*. (2012) 3:80–97. doi: 10.1002/jrsm.1037
45. Efthimiou O, Debray TP, van Valkenhoef G, Trelle S, Panayidou K, Moons KG, et al. GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods*. (2016) 7:236–63. doi: 10.1002/jrsm.1195
46. Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, et al. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value Health*. (2014) 17:157–73. doi: 10.1016/j.jval.2014.01.004
47. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med*. (2015) 162:777–84. doi: 10.7326/M14-2385
48. Hu D, O'Connor AM, Winder CB, Sargeant JM, Wang C. How to read and interpret the results of a Bayesian network meta-analysis: a short tutorial. *Anim Health Res Rev*. (2019) 20:106–15. doi: 10.1017/S1466252319000343

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

The tutorial R project with instructions, data set, scripts, bugs are available at https://github.com/a-oconnor/NETWORK_MA_FRONTIERS_TUTORIAL.

frontiers in Veterinary Science

ORIGINAL RESEARCH
published: 27 May 2020
doi: 10.3389/fvets.2020.00281

Hierarchical Structures in Livestock Trade Networks—A Stochastic Block Model of the German Cattle Trade Network

Laura Brzoska[1,2], *Mareike Fischer*[1] *and Hartmut H. K. Lentz*[2*]

[1] *Institute of Mathematics and Computer Science, University of Greifswald, Greifswald, Germany,* [2] *Institute of Epidemiology, Friedrich-Loeffler-Institut, Greifswald-Insel Riems, Greifswald, Germany*

Trade of cattle between farms forms a complex trade network. We investigate partitions of this network for cattle trade in Germany. These partitions are groups of farms with similar properties and they are inferred directly from the trade pattern between farms. We make use of a rather new method known as stochastic block modeling (SBM) in order to divide the network into smaller units. SBM turns out to outperform the more established community detection method in the context of disease control in terms of trade restriction. Moreover, SBM is also superior to geographical based trade restrictions and could be a promising approach for disease control.

Keywords: network analysis, epidemic model, cattle trade, Germany, modularity, stochastic block model

OPEN ACCESS

Edited by:
Victoria J. Brookes, Charles Sturt University, Australia

Reviewed by:
Annemarie Kaesbohrer, University of Veterinary Medicine Vienna, Austria
Gianluigi Rossi, University of Edinburgh, United Kingdom

***Correspondence:**
Hartmut H. K. Lentz
hartmut.lentz@fli.de

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *27 November 2019*
Accepted: *27 April 2020*
Published: *27 May 2020*

Citation:
Brzoska L, Fischer M and Lentz HHK (2020) Hierarchical Structures in Livestock Trade Networks—A Stochastic Block Model of the German Cattle Trade Network. Front. Vet. Sci. 7:281. doi: 10.3389/fvets.2020.00281

1. INTRODUCTION

The trade with living animals poses a major risk for the spread of infectious diseases. The latter include foot-and-mouth disease (1–3) and bovine virus diarrhea (4, 5), as well as zoonotic diseases, such as bovine tuberculosis (6, 7). Cattle farmers typically sell and/or purchase animals at a relatively high frequency and to different trading partners. Therefore, the trade between all involved farmers forms a complex network, where in case of an outbreak many farms can be infected within a short period of time.

In order to understand the structure of these trade connections as well as to quantify the risk of infection spread, the trade data can be represented as a complex network that can be analyzed mathematically (8, 9). Concerning trade data, all EU member states are obliged to report any cattle movement to a central database (10). The usage of this data is, however, typically restricted to competent authorities. Once the data is available, common network analyses focus on ranking the involved farms – nodes in the network with edges, i.e., trade connections, between them – according to their suitability for disease containment and surveillance. Highly ranked farms are then called central nodes. It has been shown that node rankings can be helpful for efficiently implementing countermeasures such as targeted vaccination (11–15). The second common goal of network analyses is understanding the large scale structure of the studied system. Typically, livestock trade networks in developed countries consist of up to 10^5 farms (15–19). Therefore, finding inherent structures that allow to partition a network into small subsets of nodes that are in the best case independent from each other, is a promising way to gain an understanding of the system as a whole. In addition, partitioning the network has another advantage: epidemics can be fought considerably better in systems consisting of smaller units. Moreover, a trade network could be constructed out of many small independent subunits on purpose. This is known as compartmentalization and can be considered as a method for passive disease protection.

The simplest example of such a partitioning is the component structure. It determines which pairs of farms can potentially infect each other at all via trade, directly or indirectly. In other words, the component structure describes whether the network consists of a large continent or a number of small disconnected islands. The component structure of a large network such as livestock trade typically yields very large structures that can be used to assign nodes to two disjoint risk classes (13, 15, 20–23). First, nodes that can reach a large number of other nodes through trade and second, nodes that can only reach a small number of others through trade.

Although partitioning nodes according to the component structure is a useful tool for risk assessment, the component structure of livestock trade networks is typically dominated by a so-called giant component (15–19, 24). That is, these networks consist of continents instead of small islands. Consequently, partitioning the network according to components does in general not yield practicable groups for disease control.

In order to find groups in networks that are applicable for disease control, the detection of so-called communities or *modules* has gained considerable attention in veterinary science in the last years (25–30). Modules are similar to components, but allow disjoint groups to be loosely connected. More precisely, a module is a group of nodes that are densely connected to each other, while they have only few connections to other modules.

Finding modules is a promising way to define compartments in networks that can in the best case be isolated from each other in case of an outbreak. Moreover, by now a number of methods is available to find modules even in large networks (31–34). Interestingly, it has been shown that in many cases the modules found for livestock trade networks also show a high spatial clustering, despite the fact that no spatial information is used to infer them [cf. Lentz et al. (26, 27, 29, 30)]. This makes modules potentially interesting for disease control (35, 36). On the other hand, it is well-known that module detection has a resolution limit, i.e., the detected modules cannot be arbitrarily small (34, 37). As an example, the modules found for pig trade in Germany have a scale of federal states (26). Therefore, partitioning such networks into modules is in most cases not feasible for disease control.

Here we use a relatively new method, Bayesian stochastic blockmodeling or simply *stochastic blockmodeling* (SBM) (38), to partition the cattle trade network in Germany into relatively small groups. The SBM method can detect smaller groups than community detection and can even find other structures than densely connected groups of nodes (38). Moreover, stochastic blockmodeling is able to find hierarchically structured groups. Therefore, we can analyze node groups of smaller sizes than those of the classical modules. In order to be applicable to disease control, these groups should (1) show geographical clustering and (2) have a resolution of at least district size, i.e., roughly 30 km (mean district diameter in Germany).

In this work, for the first time we analyze cattle trade in Germany as a complex network. We thereby put a focus on the detection of inherent groups in the network and evaluate the feasibility of different partition methods for disease control. These are community detection for finding modules, a stochastic block model, and a nested stochastic block model with hierarchical structure. Since stochastic blockmodeling is a rather novel method in veterinary applications, we also provide a detailed explanation of the method.

In order to assess the eligibility of modules and block models for animal disease control, we simulate epidemic outbreaks on the network and evaluate different control strategies based on trade restrictions according to different network partitionings. The trade restrictions are realized using targeted edge removal in the network. To compare our results to established methods for disease control, we also simulate trade restrictions based on the geographic closeness of nodes.

This article is organized as follows: We first perform a network analysis of the trade data. Then, we give an explanation of different methods for structure inference, i.e., community detection and the stochastic block model. Finally, we simulate outbreaks on the network and apply different control strategies. The results are integrated into the respective sections.

2. PROPERTIES OF THE NETWORK

2.1. Data

In this work we analyze an excerpt of the HI-Tier Database (39). The dataset contains cattle movements between farms in Germany from 2010-01-01 until 2014-12-31. Each trade item contains the source farm, target farm and the time of movement. Source and target farms are represented as *nodes* in the network and item as described above is a *trade link*. In this work, trade links are aggregated over time so that two nodes are connected by a directed *edge* whenever there is at least one trade link between them. Overall the network consists of 209,336 nodes and 1,822,373 edges. Using the trade link data without time aggregation yields a temporal network with the same number of nodes, but with 15,416,850 trade links and an observation period of 1,825 days.

2.2. Network Analysis

In this section we perform a network analysis of the cattle trade data. A summary of the network measures is given in **Table 1**. The network is represented by a graph $G = (V, E)$, where V is the set of nodes and E is the set of (directed) edges, where each edge connects a node pair.

The network can be represented by an adjacency matrix $\mathbf{A}$, where an entry $(\mathbf{A})_{ij} = 1$, if there is an edge from node i to node j, and 0 otherwise. The degree k_i of a node i is the total number of its neighbors (ingoing and outgoing), i.e., the number of its trade partners. Since we consider a directed network, we also distinguish between in-degree and out-degree for each node.

Indirect connections between nodes, traversing an arbitrary number of edges and no edges or nodes more than once, are called paths. If a path between two nodes i and j exists, we can write $i \to j$, and otherwise $i \not\to j$. A shortest path between two nodes is a path between them with a minimum number of edges. For the cattle trade network the average shortest path length is 4.4 meaning that a potential disease would take only 4.4 steps on average to infect any node in the network. The maximum shortest path length is called diameter and has a value

TABLE 1 | Properties of the static network.

Property	Value
Number of nodes	209,336
Number of edges	1,822,373
Mean degree	17.4
Mean shortest path length	4.4
Diameter	17
GWCC size	0.99
GSCC size	0.69
GIC size	0.21
GOC size	0.07
Path density	0.54

of 17 for the studied network. Considering the set of all shortest paths between all nodes in the network, the path density ρ_p is the number of such paths normalized by the number of all possible paths. The path density represents the probability that a randomly chosen node pair is connected in the network. In other words, ρ_p increases with the overall network connectivity, such that $\rho_p \to 1$ implies that all node pairs are connected via paths and $\rho_p \to 0$ implies that the network is fragmented. For our network we have $\rho_p = 0.54$.

A concept very related to paths are connected components, which are subsets of nodes $C \subset V$ such that there is a path between all node pairs in this subset. It is a well-known feature of large networks that they possess a so-called giant component, which means that the largest connected component dominates the network and is much larger than the second largest one (9). Giant components form the backbone of complex networks, since they guarantee for the most important feature: connecting nodes. Ignoring the edge directions, the resulting giant component is called giant weakly connected component (GWCC). It contains about 99% of the nodes in our network, i.e., almost all nodes are connected ignoring edge directions. If edge directions are explicitly considered, the resulting giant component is called giant strongly connected component (GSCC), and it only contains nodes that can reach each other on directed paths. The GSCC of the German cattle trade network has a size of 69% of the network nodes. Nodes that are not part of the GSCC, but can reach the latter by a path, form the giant in-component (GIC). This component consist of 21% of the network nodes. In addition, the giant out-component is formed by nodes that can be reached from the GSCC, but do not belong to it. It contains 7% of the network nodes.

3. STRUCTURE INFERENCE IN THE NETWORK

In order to efficiently implement disease control in the network based on its topology, the network has to be partitioned into groups that can be easily isolated from each other. It seems natural for this purpose to utilize components as discussed above. However, components are not practicable for disease control due to the existence of the giant component. The latter implies that most nodes belong to a single group and most other groups are irrelevant for disease spread.

On the other hand, the cattle trade network should be comprised of natural substructures—e.g., densely connected node groups or production chains. Merging these subgroups yields the observed network. They are not known from the data set and it is the aim of this section to infer these structures. We first use the well-established method of community detection, and then infer structures using the stochastic block model approach.

3.1. Community Detection

A community or module is a set of nodes, where the nodes have significantly more edges within their community than to other communities. Partitioning the network into communities in an optimal way is known to be an intractable problem for large networks (33). One way to obtain an appropriate partitioning of the network into modules is to optimize the modularity function (40, 41)

$$\begin{aligned} Q = &\text{ fraction of edges within modules} \\ &- \text{ expected fraction of these edges.} \end{aligned} \tag{1}$$

The modularity function maps a given partitioning of the network onto a single number. Optimizing Equation (1) means to find the node partition that gives the highest possible value of Q. A systematic method to find an optimal partitioning maximizing the modularity function has been proposed in Newman (25). However, the latter method is rather slow and faster methods that perform better even on larger networks have been developed (31, 42). For this work we used the *Infomap* algorithm introduced in Rosvall and Bergstrom (32), which showed a good performance in our network. It can be applied to directed networks and allows for module detection in linear time, that is, the computation time scales linearly with the number of nodes (43).

After applying the community detection algorithm to the cattle trade dataset, we find modules of sizes between 1 and 73,024. However, 99.89% of nodes are in the 10 largest modules. A map with the 10 largest modules is shown in **Figure 1**. We note that the detected modules show a high degree of spatial clustering, even though no geographical information has been used for the computation. Some of the found modules reflect borders of federal states (e.g., Rhineland-Palatinate or Hesse). Module 1 represents a whole region of Germany (Northern Germany). In addition, the modules show geographical overlap, which is more pronounced for modules 2 and 3.

3.2. Bayesian Stochastic Block Model

The idea behind a block model of a network is to find groups of nodes belonging together in some way and these groups look like dense blocks in the adjacency matrix $\mathbf{A}$ (44, 45). These blocks are also called building blocks of the network. As an example, if a network has a community structure as explained in the previous section, the adjacency matrix can be reordered such that nodes of the same module have neighboring indices (say module 1 has

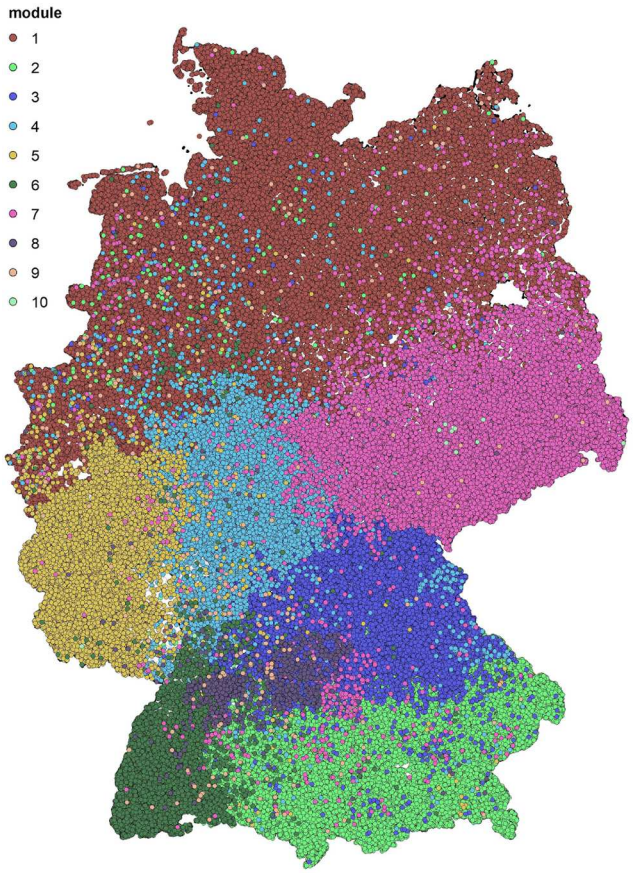


FIGURE 1 | Spatial distribution of the detected groups after modularity maximization. The ten biggest modules are shown. They show high spatial correlation.

nodes 1, . . . , 100, module 2 has nodes 101, . . . , 234, and to forth), see **Figures 2A,B**. Then, the matrix has dense blocks (many edges within communities) along the diagonal, while the rest is almost empty (few edges between different communities). The reason for this shape is that by definition node pairs of the same module have many links, while links to other modules are rare. As opposed to modules, a block model can have a more general structure, e.g., blocks far from the diagonal and no blocks on the latter. An example is shown in **Figure 2C**. Besides the fact that a block model can resolve more complex network structures than modules, dividing blocks iteratively yields a so-called hierarchical block model. As a consequence, a stochastic block model can resolve relatively small groups in a given network. In fact, it does not suffer from the resolution limit known for community detection (37).

The aim of this section is to infer the underlying block structure from a given network. We now give a brief mathematical sketch of block model inference following Peixoto (45).

At first, we consider the case where the block membership of each node is known in the first place. The network has B blocks and the block membership of the nodes is stored in a vector $\mathbf{b}$, where the entry b_i is the block membership of node i. Furthermore, the number of edges between blocks r and s is stored in a matrix $\mathbf{E}$ with entries $(\mathbf{E})_{rs}$. If we assume that nodes of the same block are statistically indistinguishable, the matrix $\mathbf{E}$ defines a set of all possible networks with the same topology. Such a set is called *ensemble*. This ensemble is the set of all virtual copies of the network with the same number of edges between blocks, i.e., the same $\mathbf{E}$. The number of nodes is also constant.

Within this ensemble, each possible network can be represented by an adjacency matrix $\mathbf{A}$. Recalling that the node partition $\mathbf{b}$ is known, the probability distribution of the possible networks is

$$P(\mathbf{A}|\mathbf{b}). \tag{2}$$

Note that this distribution is a mapping from each virtually possible network to a probability. Due to the large number of possible network configurations, Equation (2) is in general a complicated function. One way to obtain the form of the distribution Equation (2) is to maximize the entropy (or equivalently minimize the information), under certain constraints. The entropy is given by

$$S = -\sum_{\mathbf{A}} P(\mathbf{A}|\mathbf{b}) \ln P(\mathbf{A}|\mathbf{b}), \tag{3}$$

and the constraints are first, the matrix $\mathbf{E}$ containing the edges between groups, and second the normalization of the probability distribution of the networks in the ensemble. Using the method of Lagrange multipliers yields an equation for the desired probability distribution (45).

So far, we have considered the case where the node partitioning was known in the first place. Of course, the problem setting here is exactly the opposite: we have an observed network and want to infer a plausible partition of it. The central idea of the inference algorithm used here is to reverse the distribution Equation 2 using the Bayes formula

$$P(\mathbf{b}|\mathbf{A}) = \frac{P(\mathbf{A}|\mathbf{b})P(\mathbf{b})}{P(\mathbf{A})}, \tag{4}$$

where $P(\mathbf{b}|\mathbf{A})$ is the posterior distribution of network partitions given an observed network and $P(\mathbf{b})$ is the prior distribution, i.e., the distribution of network partitions in the absence of data. If we make no other assumptions, then each partition is equally likely, say $P(\mathbf{b}) = 1/N_p$, where N_p is the number of possible partitions. The term $P(\mathbf{A}|\mathbf{b})$ is called evidence and describes the impact of the network data on the prior information, and $P(\mathbf{A})$ is a normalization constant.

This way we obtain a probability distribution of network partitions ($\mathbf{b}$) given an observed network ($\mathbf{A}$). The task of finding the optimal partition is equivalent to finding the partition maximizing the posterior distribution, i.e., the left-hand side of Equation (4).

Although formal solutions for this equation exist, these solutions are too complex to find their maxima or even sample from them. Note that, similar to the modularity function Equation (1), Equation (4) maps each possible partition onto

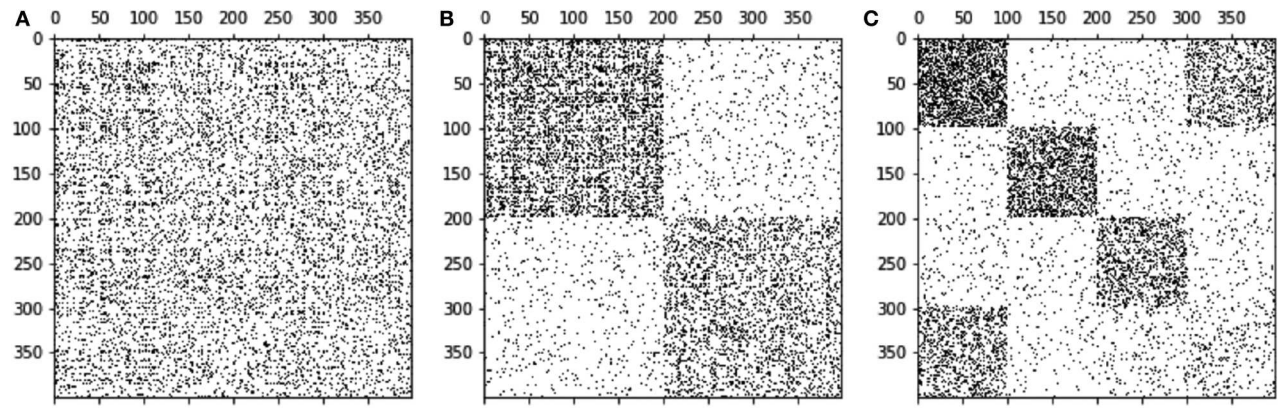


FIGURE 2 | Three different representations of the same adjacency matrix. Through ordering the nodes in various ways we obtain different patterns. Panel **(A)** displays a random order of nodes. Panel **(B)** shows a modular structure, whereas in **(C)** a block structure is observed.

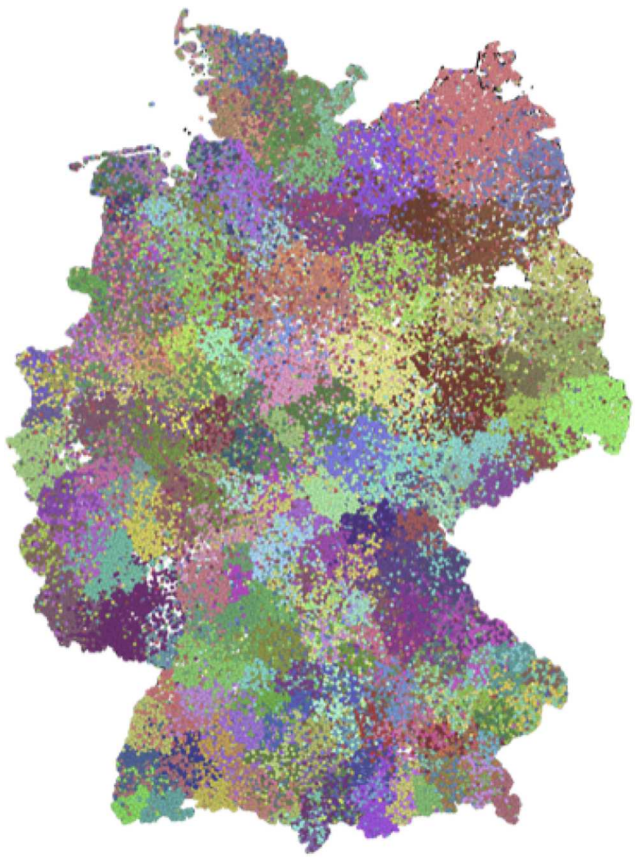


FIGURE 3 | Partition after a block model with 382 blocks. The sizes of blocks roughly correspond to sizes of districts.

TABLE 2 | Block sizes of the hierarchical model.

Level	Number of blocks
j	1
i	2
h	4
g	12
f	39
e	84
d	180
c	370
b	974
a	2,956

The block size is measured as the number of blocks in the subjacent level.

a probability. The combinatorial number of such partitions N_p (known as the bell number) is extremely large, and finding the optimal partition of a network is an NP-hard problem, i.e., it is intractable for large networks. For this reason, we utilize a Markov Chain Monte Carlo (MCMC) approach. The idea behind this approach is to start with an arbitrary partition $\mathbf{b}_0$ and change this partition to $\mathbf{b}_1$. This can be realized changing the group membership of a single node. Such a change is accepted, if the resulting partition $\mathbf{b}_1$ increases the posterior distribution (left-hand side of Equation 4). Even if it decreases the latter, it is still accepted with a certain probability. In the long term this procedure results in a random walk in the space of possible partitions and defines a way to sample from the posterior distribution. For the above procedure to converge to the maximum of the distribution, one can slowly reduce the mobility of the random walk, until it remains at the most probable position. This is known as simulated annealing (46).

Although the algorithm above is applicable to the partition problem, it has been shown that convergence can be slow on large networks. For this reason, an optimized MCMC method has been proposed in Peixoto (47). This method is a greedy agglomerative heuristic, i.e., we start with each node being one block and then group nodes together successively. In contrast, in a divisive algorithm, one would start with the whole network as one block and divide until only nodes are left. Divisive algorithms, however, are computationally expensive and agglomerative algorithms are commonly used for large datasets.

The purpose of this section is to give an intuitive understanding of the method of inferring block structures. For

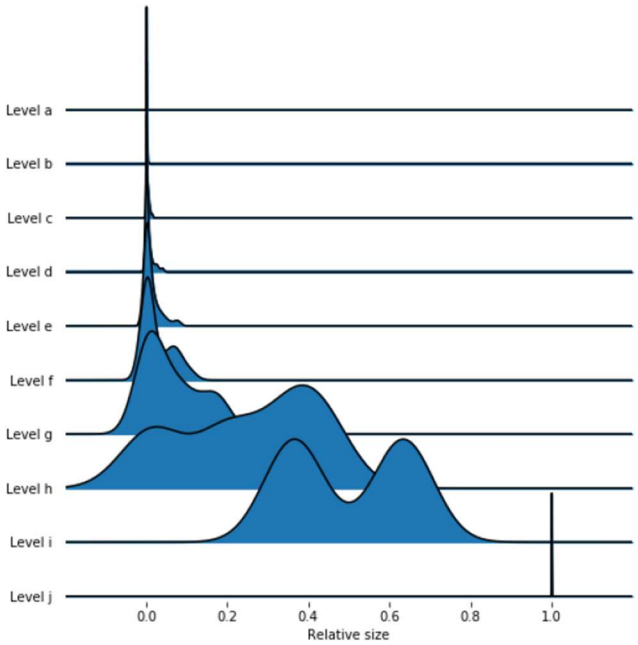


FIGURE 4 | Distributions of block sizes on the different levels. The lower levels have smaller block sizes, while the upper levels are larger.

a far more comprehensive explanation the reader is referred to Peixoto (38, 45). An implementation of block model inference including the hierarchical version is provided in the software package *graph-tool* (48).

Using this software package, we infer a block partition of the German cattle trade network and obtain a partition into 382 blocks. The sizes of blocks are between 1 and 2,887. A map with the block membership of the farms is shown in **Figure 3**. Similar to the modules, we note that the found blocks show a high degree of spatial clustering, although no geographical information has been used to compute them. Some of the detected blocks reflect borders of districts. Furthermore, some blocks show a geographical overlap.

In addition to this partitioning technique, a block model partition can be inferred for each detected block iteratively (38). The result is called hierarchical or *nested* model. It has the following properties: First, the outcome is a block model where each block is divided into smaller sub-blocks. Second, the resolution of the detected blocks can be increased this way, i.e., blocks found in the lowest hierarchy should have a smaller size than the blocks found using the non-hierarchical method.

Using a *nested* block model to resolve the hierarchical structure of the network, we obtain a hierarchy of ten levels, which are labeled from *a* to *j*. **Table 2** contains the number of blocks in each level, while **Figure 4** displays the distributions of the block sizes. We consider the highest three levels *g*, *h* and *i* in the hierarchy. Level *j* represents the whole country and is therefore trivial. **Figures 5A–C** shows the various levels. As above the blocks show a high geographical correlation. The blocks of level *i* strictly divide the country into north (red) and south (blue) Germany. The subjacent level *h* divides the farms by the borders of the regions north (red), central (yellow), and south Germany (blue). Blocks of level *g* (**Figure 5C**) still reflect larger geographical regions. The north-eastern block (light green points in **Figure 5C**) contains large parts of federal states like Mecklenburg-Western Pomerania, Schleswig-Holstein, Brandenburg and parts of Lower Saxony. Some blocks reflect borders of federal states (e.g., Bavaria, Baden-Wuerttemberg) (blue points in **Figure 5C**). The red block spans over several federal states (e.g., Lower Saxony and North Rhine-Westphalia and others). It also shows a large geographical overlap with the light green block. **Figure 6** shows the trade links (network edges) between all blocks. The hierarchy is represented by the blue tree, where the root node represents level *j* (whole Germany), its neighbors are level *i* and so on. In the center the dominant branch separating northern from southern Germany is clearly visible. Although the trade structure appears to be complicated, **Figure 6** demonstrates that trade links are distributed rather homogeneously between the blocks.

Finally, we check to what extent the nested block model gives a similar result as the non-nested version. Therefore, we choose level *c* (**Figure 7**), since the number of blocks here is similar to the non-nested version. The figure shows qualitative similarities to the non-nested model (**Figure 3**). In addition, the SBM on level *c* has similarities with the detected modules (see **Supplementary Material**).

4. USING THE INFERRED STRUCTURES FOR DISEASE CONTROL

We evaluate the applicability of the detected modules and block models on disease control by simulating epidemics on the network. Thereby, we utilize different control strategies (see below) based on trade restrictions and compare them to established methods. The established control strategy is based on geographical trade restrictions around a certain radius around the farm where a disease was detected.

The infection process is modeled using a so called SI-model, where an infected farm (I) contaminates a susceptible farm (S) upon trade contact with a rate β. Once a farm i is infected, it can infect its neighbors, i.e., farms being connected to i by a trade link, with rate β. An infected farm stays infectious during the whole simulation. In order to guarantee stable results, we have to choose different initial conditions. Thus, we first sample 10 blocks and then sample 10 nodes out of each block as starting nodes. If we modeled the epidemic process as explained, the disease would infect large fractions of the network within a few steps, since the network is static and all infectious links are permanently active. Therefore, we mimic the temporal nature of trade considering only a fraction of network edges as being present at each step. This fraction can be estimated considering the time span, where a farm does not trade. Out of the data, we observe that nodes are only active every 10 days on average. **Figure 8** shows the waiting time distribution (the time span where a farm does not trade) for

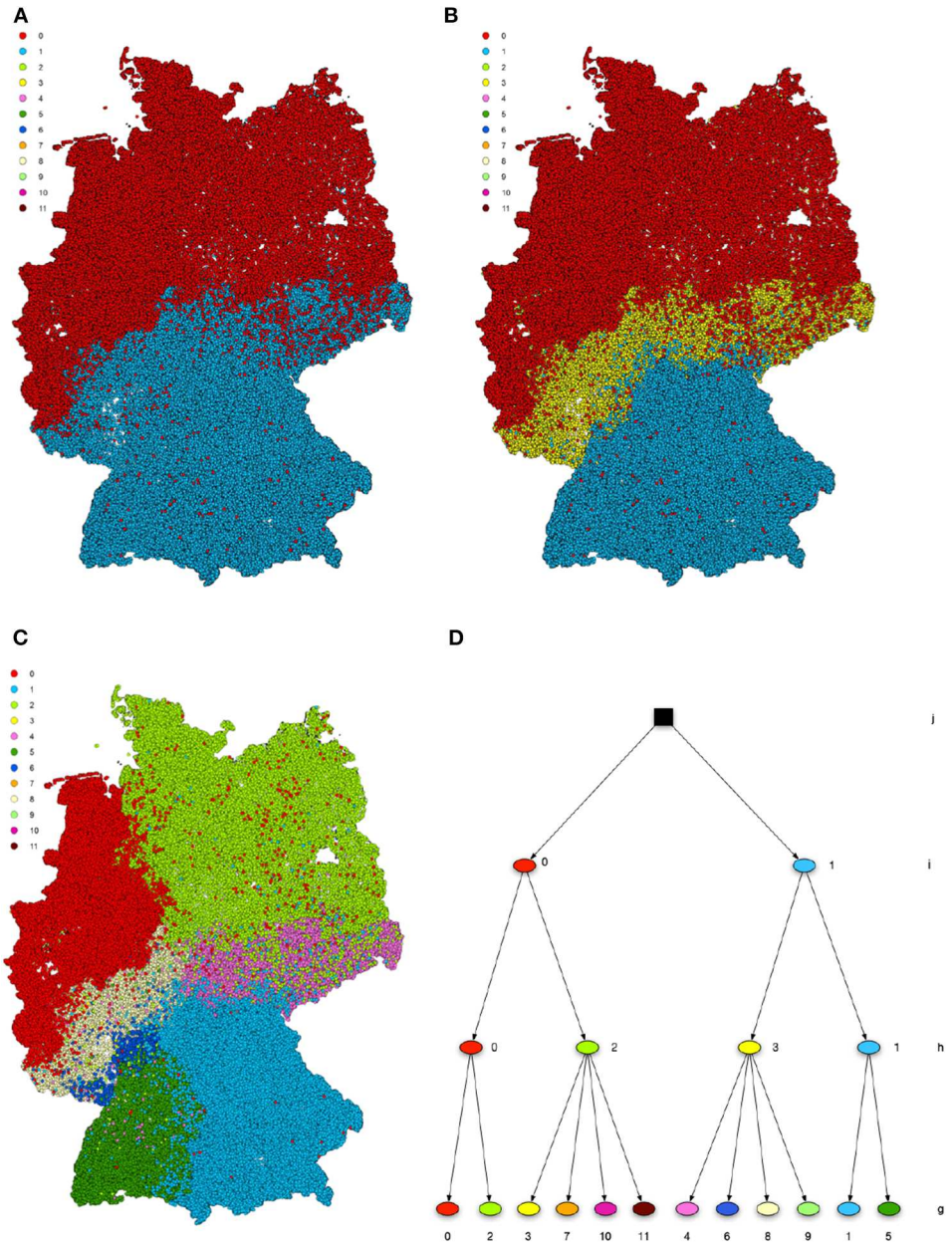


FIGURE 5 | Hierarchical structure after a nested SBM. The levels *i* **(A)**, *h* **(B)**, and *g* **(C)** are geographically shown. The blocks show high spatial correlation on all levels of the hierarchy. Panel **(D)** shows the hierarchy tree.

all nodes of the network. Since the mean value of the waiting time distribution is around 10, we mimic the waiting times using the infection rate, and set $\beta = 0.1$.

In order to assess the performance of different control strategies on the simulated outbreak, we simulate SI dynamics for the different starting nodes and compare the results of the non-nested and nested stochastic block model (level *c*) to the modules found with community detection and the geographical method as explained above. We thereby model disease control measures in terms of trade restrictions. These are realized by removing edges of the trade network according to different schemes. The first control strategy is based on geographical trade restrictions around a 10 km radius of the index farm after detection of the disease (49). Second, we evaluate the applicability of the found modules and blocks by applying trade restrictions at the (edge) boundary of the respective structures. That is, we remove all edges of the module/block of the index farm that point to other modules/blocks, respectively. The different strategies are shown in **Figure 9**.

Assuming that a disease will spread freely only before it is detected, we remove the trade contacts regarding the infection start node after a detection time t_d. We choose $t_d = 1$ day.

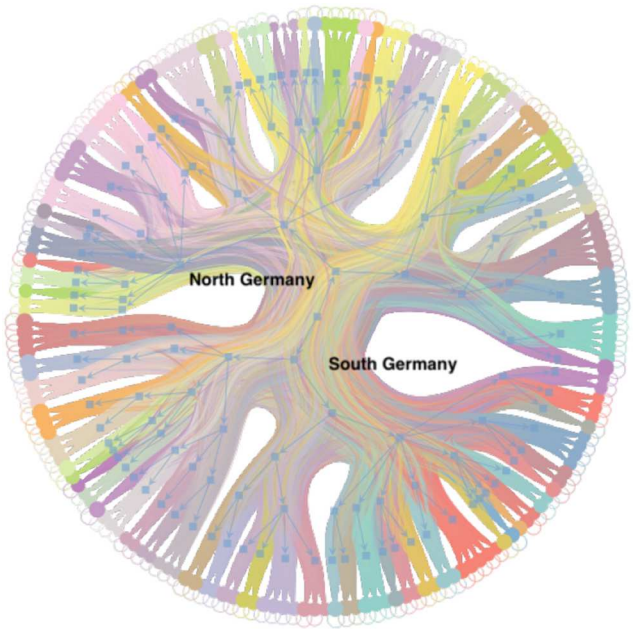


FIGURE 6 | Hierarchy of trade links between blocks of **Figure 5**. The outer nodes correspond to blocks on level *d*. The blue tree shows the hierarchy and each node of the tree corresponds to a block. The dominant branch in the center separates the northern from the southern German blocks.

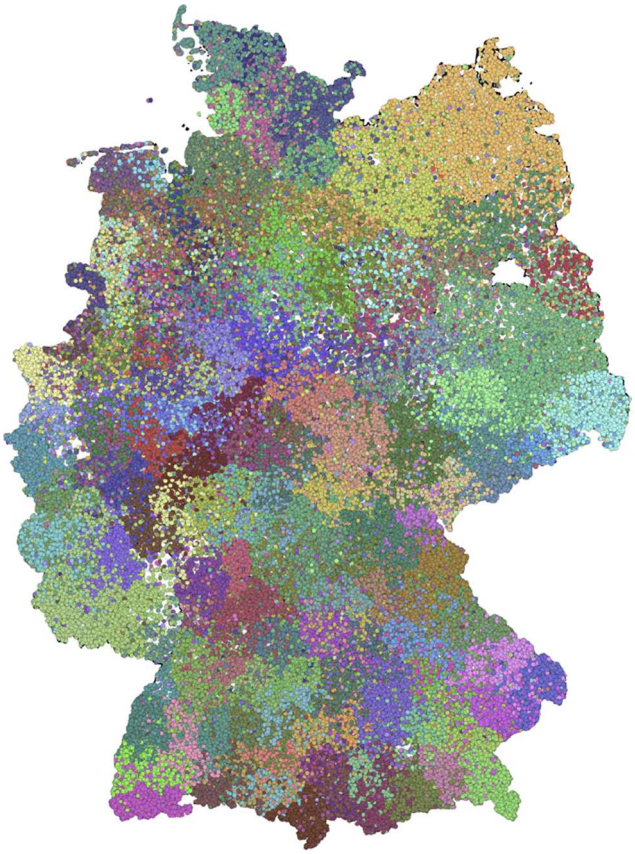


FIGURE 7 | Spatial distribution of level *c* of the nested block model. This level has a block size similar to the non-nested case.

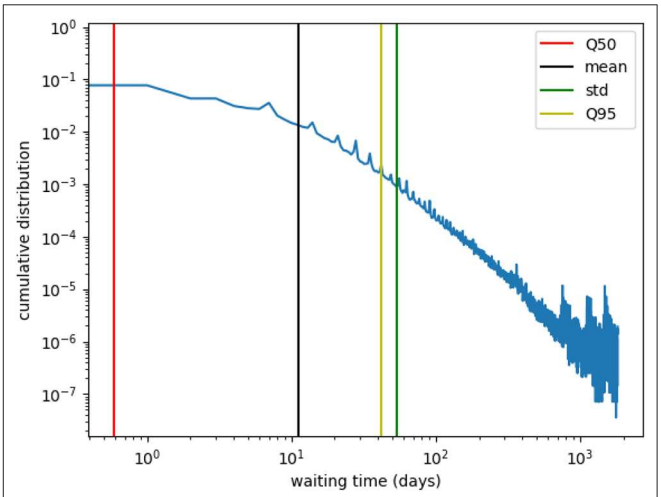


FIGURE 8 | Farm waiting time distribution. The waiting time is the interval in which a farm is not active, e.g., it does not trade. The mean value is roughly 10, i.e., farms trade every 10-th day on average. Therefore, we choose an infection rate $\beta = 0.1$.

This can be considered a best case scenario. Indeed, the detection time depends on the incubation period of the considered disease. However, choosing other values for the detection time does not change the results qualitatively (see **Supplementary Material**). Due to the fact that the infection process is stochastic, we run the simulation ten times for each starting node.

Figure 10 shows the results of the different control strategies for the nested and non-nested block model. We determine for each strategy the mean values of the outbreak size and the number of edges, which were removed in the simulation. As a consequence of this result the block model leads to a slightly higher number of removed edges in comparison to the geographical method. However, the outbreak size of the block model is significantly smaller. Both strategies, the non-nested block model and the geographical method, perform better than module based trade restrictions in the sense that many edges have to be removed for the latter case and the outbreak sizes are still relatively large. This is due to the fact that the size of the modules is significantly larger than the sizes of the other two groups. **Figure 11** shows the sizes of the different groups on a map.

Concerning the nested block model (level *c*), the results are very similar to the non-nested case (see **Figure 10**). The only difference here is that slightly more edges have to be removed for trade restriction.

An evaluation of different values for the parameters detection time d, infection rate β, and radius around the index farm for trade restrictions, is provided in the **Supplementary Material**. All parameters affect the outbreak sizes and number of removed edges systematically, but do not alter the qualitative results of **Figure 10**.

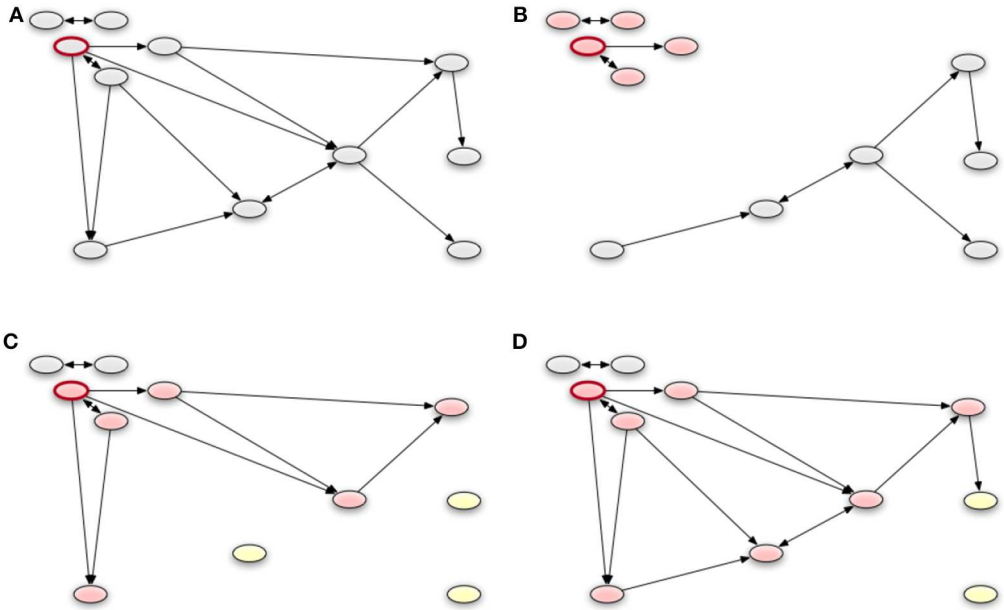


FIGURE 9 | An example of applying the different control strategies of a network. The red framed node represents the index node. We remove all edges of the module or block of the index node that point to other modules or blocks, respectively. Panel **(A)** shows the whole the network. The geographical strategy is represented in **(B)**. Nodes which have a small distance to the index node, build a group (red colored) and there are no edges to nodes with a long distance to the index node (gray). In **(C)**, we see a possible result of the stochastic block model with the blocks. All edges from the block of the index node to the other blocks are removed. A similar result is shown in **(D)**, which presents the groups of modularity optimization. It should be noted that there the group is still similar to the corresponding block but contains more nodes.

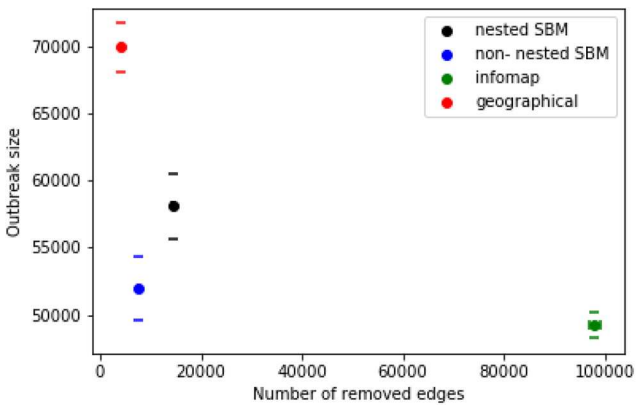


FIGURE 10 | Outbreak size and number of removed edges for different control strategies. Geographic based disease control (red) compared with control based on the block model. Trade restrictions isolating the block of the nested block model of the index node are shown in black, trade restrictions isolating the module of the index node are green. Furthermore the average error of the mean value is shown for each result.

In summary, geographically based trade restrictions are superior to module based restrictions. However, stochastic block model based trade restriction outperforms the geographical method yielding smaller outbreaks at a similar number of removed edges.

5. DISCUSSION

In this work, we have analyzed the cattle trade network in Germany for the first time. The focus was to evaluate the applicability of different partitions of the network for disease control. As a relatively new method for network partitioning we have used a stochastic block model to infer densely connected farm groups. In contrast to the well-established community detection algorithms, the stochastic block model is capable of detecting relatively small farm groups and can even be used to infer a hierarchical structure. We have found that applying trade restrictions based on a stochastic block model is more efficient for disease control than geographical, or module based trade restrictions.

Disease control has been implemented in this work as trade restrictions, and the disease spread follows a relatively simple model, i.e., the SI-model. Even though this model oversimplifies the course of most relevant diseases, the infection mechanism in the beginning of the outbreak can be approximated by an SI-process in most cases. We provide a comparison between the SI-model and the SIR-model (susceptible - infected - recovered), where farms are removed from the infection process after a certain period, in the **Supplementary Material**. The difference between the two models is marginal for detection times less than 14 days.

In contrast to geographically based trade restrictions network based restrictions are not guaranteed to be constant over time since trade patterns might change (15, 18, 24, 50, 51). In

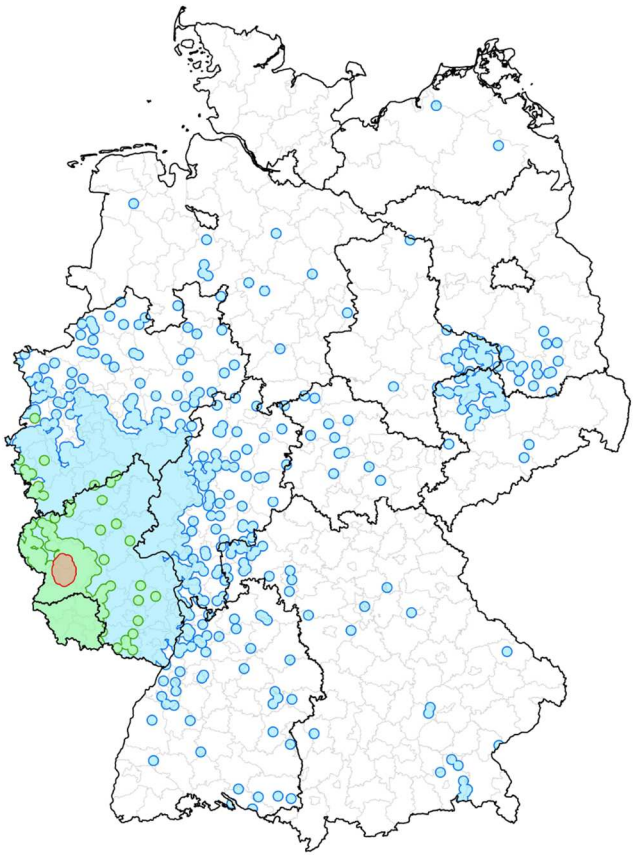


FIGURE 11 | An example to illustrate the different memberships of the same farm. The red area describes the ten kilometer radius around the chosen node and represents the geographical control strategy. Furthermore, the green colored area represents the block affiliation. The module membership of the node is shown as the blue surface.

applications trade boundaries could simply be computed using current trade data so that temporal constancy plays a minor role. Moreover, it is plausible that particularly larger node groups show only small fluctuations over time (24).

As we have demonstrated, modules as well as blocks show a high degree of spatial clustering. Even though this property is also used in the geographical approach, the trade data offers still another way of node partition: the underlying production chains. A stochastic block model should in principle be capable of finding such structures as well. For example, functional blocks in the world trade network have been found in Reichardt and White (52), where the authors could resolve the role of different countries in global economy. However, this requires a relatively complex null model (mathematically speaking in the form of constraints in the optimization) in the inference algorithm. It would be interesting for future work to validate different null models in order to resolve production chains. If the production chains were known, we could implement economically efficient trade restrictions allowing for redirecting trade channels in the case of an outbreak.

As our results show, the application of a hierarchical block model on cattle trade data seems to be a promising approach for applications in livestock disease control. Moreover, decoupling trade restrictions from geographical neighborhood protects the neighborhood of the index farm from being considered false positive, and thus might contribute to animal welfare. However, these statements only holds if we neglect current legislation for disease control, and it is of course beyond the scope of this paper to change legislation. Nevertheless, the strategy for trade restriction presented here is technically feasible, i.e., only low computational power is needed and block structures could be inferred on the fly, or at least on a regular basis, in order to have on-time trade groups. We therefore believe that groups in trade data are useful in application and could improve disease control.

DATA AVAILABILITY STATEMENT

The datasets analyzed in this article are not publicly available. Requests to access the datasets should be directed to hartmut.lentz@fli.de.

AUTHOR CONTRIBUTIONS

LB performed the analyses. LB and HL wrote the manuscript. HL and MF designed the study.

ACKNOWLEDGMENTS

We thank Dr. Nicolai Denzin for fruitful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.00281/full#supplementary-material

REFERENCES

1. Bouma A, Elbers A, Dekker A, De Koeijer A, Bartels C, Vellema P, et al. The foot-and-mouth disease epidemic in the Netherlands in 2001. *Prev Vet Med*. (2003) 57:155–66. doi: 10.1016/S0167-5877(02)00217-9
2. Pluimers F, Akkerman A, Van Der Wal P, Dekker A, Bianchi A. Lessons from the foot and mouth disease outbreak in the Netherlands in 2001. *Rev Sci Tech OIE*. (2002) 21:711–8. doi: 10.20506/rst.21.3.1371
3. Martínez-López B, Perez A, De la Torre A, Rodriguez JSV. Quantitative risk assessment of foot-and-mouth disease introduction into Spain via importation of live animals. *Prev Vet Med*. (2008) 86:43–56. doi: 10.1016/j.prevetmed.2008.03.003
4. Meyling A, Houe H, Jensen A. Epidemiology of bovine virus diarrhoea virus. *Rev Sci Tech OIE*. (1990) 9:75–93. doi: 10.20506/rst.9.1.489
5. Gethmann J, Homeier T, Holsteg M, Schirrmeier H, Saßerath M, Hoffmann B, et al. BVD-2 outbreak leads to high losses in cattle farms in Western Germany. *Heliyon*. (2015) 1:e00019. doi: 10.1016/j.heliyon.2015.e00019
6. Schiller I, RayWaters W, Vordermeier HM, Jemmi T, Welsh M, Keck N, et al. Bovine tuberculosis in Europe from the perspective

of an officially tuberculosis free country: trade, surveillance and diagnostics. *Vet Microbiol*. (2011) 151:153–9. doi: 10.1016/j.vetmic.2011.02.039
7. Probst C, Freuling C, Moser I, Geue L, Köhler H, Conraths F, et al. Bovine tuberculosis: making a case for effective surveillance. *Epidemiol Infect*. (2011) 139:105–12. doi: 10.1017/S0950268810000786
8. Martínez-López B, Perez AM, Sánchez-Vizcaíno JM. Social network analysis. Review of general concepts and use in preventive veterinary medicine. *Transbound Emerg Dis*. (2009) 56:109–20. doi: 10.1111/j.1865-1682.2009.01073.x
9. Newman MEJ. The structure and function of complex networks. *SIAM Rev*. (2003) 45:167–256. doi: 10.1137/S003614450342480
10. EUR-Lex. Directive 2000/15/EC of the European Parliament and the Council of 10 April 2000 Amending Council Directive 64/432/EEC on Health Problems Affecting Intra-Community Trade in Bovine Animals and Swine (2000). Available online at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32000L0015
11. Albert R, Jeong H, Barabási AL. Error and attack tolerance of complex networks. *Nature*. (2000) 406:378–82. doi: 10.1038/35019019
12. Holme P, Kim BJ, Yoon CN, Han SK. Attack vulnerability of complex networks. *Phys Rev E*. (2002) 65:056109. doi: 10.1103/PhysRevE.65.056109
13. Natale F, Savini L, Giovannini A, Calistri P, Candeloro L, Fiore G. Evaluation of risk and vulnerability using a disease flow centrality measure in dynamic cattle trade networks. *Prev Vet Med*. (2011) 98:111–8. doi: 10.1016/j.prevetmed.2010.11.013
14. Büttner K, Krieter J, Traulsen A, Traulsen I. Efficient interruption of infection chains by targeted removal of central holdings in an animal trade network. *PLoS ONE*. (2013) 8:e74292. doi: 10.1371/journal.pone.0074292
15. Lentz HHK, Koher A, Hövel P, Gethmann J, Sauter-Louis C, Selhorst T, et al. Disease spread through animal movements: a static and temporal network analysis of pig trade in Germany. *PLoS ONE*. (2016) 11:e0155196. doi: 10.1371/journal.pone.0155196
16. Christley R, Robinson S, Lysons R, French N. Network analysis of cattle movement in Great Britain. In: *Proceedings: Society for Veterinary Epidemiology and Preventive Medicine*. Edinburgh: SVEPM. (2005) 234–43.
17. Bajardi P, Barrat A, Savini L, Colizza V. Optimizing surveillance for livestock disease spreading through animal movements. *J R Soc Interface*. (2012) 9:2814–25. doi: 10.1098/rsif.2012.0289
18. Dutta BL, Ezanno P, Vergu E. Characteristics of the spatio-temporal network of cattle movements in France over a 5-year period. *Prev Vet Med*. (2014) 117:79–94. doi: 10.1016/j.prevetmed.2014.09.005
19. Palisson A, Courcoul A, Durand B. Analysis of the spatial organization of pastures as a contact network, implications for potential disease spread and biosecurity in livestock, France, 2010. *PLoS ONE*. (2017) 12:e0169881. doi: 10.1371/journal.pone.0169881
20. Dorogovtsev S, Mendes J, Samukhin A. Giant strongly connected component of directed networks. *Phys Rev E*. (2001) 64:025101. doi: 10.1103/PhysRevE.64.025101
21. Dubé C, Ribble C, Kelton D, McNab B. Comparing network analysis measures to determine potential epidemic size of highly contagious exotic diseases in fragmented monthly networks of dairy cattle movements in Ontario, Canada. *Transbound Emerg Dis*. (2008) 55:382–392. doi: 10.1111/j.1865-1682.2008.01053.x
22. Nöremark M, Håkansson N, Lewerin SS, Lindberg A, Jonsson A. Network analysis of cattle and pig movements in Sweden: measures relevant for disease control and risk based surveillance. *Prev Vet Med*. (2011) 99:78–90. doi: 10.1016/j.prevetmed.2010.12.009
23. Lentz HHK, Selhorst T, Sokolov IM. Spread of infectious diseases in directed and modular metapopulation networks. *Phys Rev E*. (2012) 85:066111. doi: 10.1103/PhysRevE.85.066111
24. Knific T, Ocepek M, Kirbiš A, Lentz HHK. Implications of cattle trade for the spread and control of infectious diseases in Slovenia. *Front Vet Sci*. (2020) 6:454. doi: 10.3389/fvets.2019.00454
25. Newman MEJ. Modularity and community structure in networks. *Proc Natl Acad Sci USA*. (2006) 103:8577–82. doi: 10.1073/pnas.0601602103
26. Lentz HHK, Konschake M, Teske K, Kasper M, Rother B, Carmanns R, et al. Trade communities and their spatial patterns in the German pork production network. *Prev Vet Med*. (2011) 98:176–81. doi: 10.1016/j.prevetmed.2010.10.011
27. Rautureau S, Dufour B, Durand B. Structural vulnerability of the French swine industry trade network to the spread of infectious diseases. *Animal*. (2012) 6:1152–62. doi: 10.1017/S1751731111002631
28. Grisi-Filho JHH, Amaku M, Ferreira F, Dias RA, Neto JSF, Negreiros RL, et al. Detecting livestock production zones. *Prev Vet Med*. (2013) 110:304–11. doi: 10.1016/j.prevetmed.2012.12.013
29. Gorsich EE, Luis AD, Buhnerkempe MG, Grear DA, Portacci K, Miller RS, et al. Mapping U.S. cattle shipment networks: Spatial and temporal patterns of trade communities from 2009 to 2011. *Prev Vet Med*. (2016) 134:82–91. doi: 10.1016/j.prevetmed.2016.09.023
30. Relun A, Grosbois V, Sánchez-Vizcaíno JM, Alexandrov T, Feliziani F, Waret-Szkuta A, et al. Spatial and functional organization of pig trade in different european production systems: implications for disease prevention and control. *Front Vet Sci*. (2016) 3:4. doi: 10.3389/fvets.2016.00004
31. Clauset A, Newman MEJ, Moore C. Finding community structure in very large networks. *Phys Rev E*. (2004) 70:066111. doi: 10.1103/PhysRevE.70.066111
32. Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci USA*. (2008) 105:1118–23. doi: 10.1073/pnas.0706851105
33. Fortunato S. Community detection in graphs. *Phys Rep*. (2010) 486:75–174. doi: 10.1016/j.physrep.2009.11.002
34. Good BH, de Montjoye YA, Clauset A. Performance of modularity maximization in practical contexts. *Phys Rev E*. (2010) 81:046106. doi: 10.1103/PhysRevE.81.046106
35. Masuda N. Immunization of networks with community structure. *New J Phys*. (2009) 11:123018. doi: 10.1088/1367-2630/11/12/123018
36. Salathé M, Jones JH. Dynamics and control of diseases in networks with community structure. *PLOS Comput Biol*. (2010) 6:e1000736. doi: 10.1371/journal.pcbi.1000736
37. Fortunato S, Barthélemy M. Resolution limit in community detection. *Proc Natl Acad Sci USA*. (2007) 104:36–41. doi: 10.1073/pnas.0605965104
38. Peixoto TP. Hierarchical block structures and high-resolution model selection in large networks. *Phys Rev X*. (2014) 4:011047–18. doi: 10.1103/PhysRevX.4.011047
39. StMELF. Herkunftssicherungs und Informationssystem für Tiere (2012). Available online at: http://hi-tier.de (accessed August 10, 2019).
40. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Phys Rev E*. (2004) 69:026113. doi: 10.1103/PhysRevE.69.026113
41. Leicht EA, Newman MEJ. Community structure in directed networks. *Phys Rev Lett*. (2008) 100:118703. doi: 10.1103/PhysRevLett.100.118703
42. Newman MEJ. Fast algorithm for detecting community structure in networks. *Phys Rev E*. (2004) 69:066133. doi: 10.1103/PhysRevE.69.066133
43. Yang Z, Algesheimer R, Tessone CJ. A Comparative analysis of community detection algorithms on artificial networks. *Sci Rep*. (2016) 6:30750. doi: 10.1038/srep30750
44. Holland PW, Laskey KB, Leinhardt S. Stochastic blockmodels: first steps. *Soc Netw*. (1983) 5:109–37. doi: 10.1016/0378-8733(83)90021-7
45. Peixoto TP. Bayesian stochastic blockmodeling. In: *Advances in Network Clustering and Blockmodeling*. Doreian P, Batagelj V, Ferligoj A, editors. New York, NY: John Wiley & Sons, Ltd. (2019). p. 289–332. doi: 10.1002/9781119483298.ch11
46. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. *Science*. (1983) 220:671–80. doi: 10.1126/science.220.4598.671

47. Peixoto TP. Efficient Monte Carlo and greedy heuristic for the inference of stochastic block models. *Phys Rev E*. (2014) 89:012804. doi: 10.1103/PhysRevE.89.012804
48. Peixoto TP. The Graph-Tool Python Library. figshare (2014). Available online at: http://figshare.com/articles/graph_tool/1164194 (accessed October 25, 2019).
49. Dataset of the EU, C. (2003). Eu legislation egrl 85/2003 (celex nr: 32003l0085). Available online at: https://eur-lex.europa.eu/legal-content/DE/ALL/?uri=CELEX%3A32003L0085
50. Widgren S, Frössling J. Spatio-temporal evaluation of cattle trade in Sweden: description of a grid network visualization technique. *Geospatial Health*. (2010) 5:119–130. doi: 10.4081/gh.2010.192
51. Bajardi P, Barrat A, Natale F, Savini L, Colizza V. Dynamical patterns of cattle trade movements. *PLoS ONE*. (2011) 6:e19869. doi: 10.1371/journal.pone.0019869
52. Reichardt J, White D. Role models for complex networks. *Eur Phys J*. (2007) 60:217–24. doi: 10.1140/epjb/e2007-000340-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

REVIEW
published: 26 June 2020
doi: 10.3389/fvets.2020.00314

Critically Appraised Topics (CATs) in Veterinary Medicine: Applying Evidence in Clinical Practice

Marnie L. Brennan[1], Sebastian P. Arlt[2], Zoe Belshaw[3], Louise Buckley[4], Louise Corah[5], Hannah Doit[1], Virginia R. Fajt[6], Douglas J. C. Grindlay[7], Heather K. Moberly[8], Lisa D. Morrow[5], Jenny Stavisky[5] and Constance White[9]*

[1] Centre for Evidence-based Veterinary Medicine, University of Nottingham, Loughborough, United Kingdom, [2] Clinic for Animal Reproduction, Faculty of Veterinary Medicine, Freie Universität Berlin, Berlin, Germany, [3] PDSA Nottingham, Nottingham, United Kingdom, [4] Deanery of Clinical Sciences, College of Medicine & Veterinary Medicine, University of Edinburgh, Edinburgh, United Kingdom, [5] School of Veterinary Medicine and Science, University of Nottingham, Loughborough, United Kingdom, [6] Veterinary Physiology and Pharmacology, Texas A&M University College of Veterinary Medicine and Biomedical Sciences, College Station, TX, United States, [7] Centre of Evidence Based Dermatology, University of Nottingham, Nottingham, United Kingdom, [8] Medical Sciences Library, University Libraries, Texas A&M University, College Station, TX, United States, [9] Fremont Veterinary Clinic, Portland, OR, United States

OPEN ACCESS

Edited by:
Victoria J. Brookes, Charles Sturt University, Australia

Reviewed by:
Hsin-Yi Weng, Purdue University, United States
Caitlin Pfeiffer, The University of Melbourne, Australia

***Correspondence:**
Marnie L. Brennan
marnie.brennan@nottingham.ac.uk

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *31 January 2020*
Accepted: *06 May 2020*
Published: *26 June 2020*

Citation:
Brennan ML, Arlt SP, Belshaw Z, Buckley L, Corah L, Doit H, Fajt VR, Grindlay DJC, Moberly HK, Morrow LD, Stavisky J and White C (2020) Critically Appraised Topics (CATs) in Veterinary Medicine: Applying Evidence in Clinical Practice. Front. Vet. Sci. 7:314. doi: 10.3389/fvets.2020.00314

Critically appraised topics (CATs) are evidence syntheses that provide veterinary professionals with information to rapidly address clinical questions and support the practice of evidence-based veterinary medicine (EBVM). They also have an important role to play in both undergraduate and post-registration education of veterinary professionals, in research and knowledge gap identification, literature scoping, preparing research grants and informing policy. CATs are not without limitations, the primary one relating to the rapid approach used which may lead to selection bias or restrict information identified or retrieved. Furthermore, the narrow focus of CATs may limit applicability of the evidence findings beyond a specific clinical scenario, and infrequently updated CATs may become redundant. Despite these limitations, CATs are fundamental to EBVM in the veterinary profession. Using the example of a dog with osteoarthritis, the five steps involved in creating and applying a CAT to clinical practice are outlined, with an emphasis on clinical relevance and practicalities. Finally, potential future developments for CATs and their role in EBVM, and the education of veterinary professionals are discussed. This review is focused on critically appraised topics (CATs) as a form of evidence synthesis in veterinary medicine. It aims to be a primary guide for veterinarians, from students to clinicians, and for veterinary nurses and technicians (hereafter collectively called veterinary professionals). Additionally, this review provides further information for those with some experience of CATs who would like to better understand the historic context and process, including further detail on more advanced concepts. This more detailed information will appear in pop-out boxes with a double-lined surround to distinguish it from the information core to producing and interpreting CATs, and from the boxes with a single line surround which contain additional resources relevant to the different parts of the review.

Keywords: critically appraised topic (CAT), knowledge summary, BestBETs, evidence synthesis, evidence-based veterinary medicine, veterinary medicine, clinical practice

EVIDENCE-BASED VETERINARY MEDICINE

Evidence-based veterinary medicine (EBVM) can be defined as the application of scientifically generated evidence into clinical veterinary practice, whilst synergistically incorporating the expertise of the veterinary professional, the specific features of the patient and the values of the owner (1). In order to practice EBVM, it is important for veterinary professionals to keep up to date with the latest research findings to ensure they are providing the best possible care for patients they treat (2). This is challenging due to the vast amount of information published every day, and for professionals working in the current framework of veterinary practice, it is difficult to find the time (3). Additionally, it can be challenging to interpret the published literature to determine whether it is of relevance, to identify whether the results of the studies are valid and the conclusions drawn by the authors appropriate (4). Structured summaries of the published research (evidence syntheses) are of huge benefit to veterinary professionals, allowing them to easily and quickly incorporate evidence into clinical practice.

EVIDENCE SYNTHESES: REVIEWS AND CRITICALLY APPRAISED TOPICS

Most people will have heard of "literature reviews" or "narrative reviews." They are typically written by experts who summarize a number of information sources, often peer reviewed articles, on a particular area of interest and offer conclusions. They rarely control for bias or follow a specific methodology for identifying and selecting the sources that are included. Without these standards, the review may not cover the topic inclusively and the conclusions may support a specific agenda or view.

Evidence syntheses [also known as "research syntheses" or "knowledge syntheses" (5)] collectively describe a range of approaches for more objectively summarizing the literature (6). Methodological differences between types of evidence syntheses include the processes and standards for identifying, selecting, and analyzing the sources reviewed and included (7). These methodological variations support differences in the efforts to control for bias, size of the project team, comprehensiveness, and duration. Systematic reviews (SRs) are a type of evidence synthesis that follow a structured methodology to ensure all the available evidence (published and unpublished) is identified and considered (8).

Critically appraised topics (CATs) use the principles of SRs to minimize bias in gathering and appraising evidence, but do so much more quickly (5, 9–11). A CAT is based on a question of interest originating from professionals asking the question after an encounter with a particular clinical case or situation (12).

Evidence synthesis methods exist along a spectrum of brevity and detail; CATs are the quickest, SRs the lengthiest and most thorough, and other types fall in between (13). As well as speed and detail, the scope of the question, qualifications of the reviewer and the risk of bias may also differ between the different types of review (3) (**Figure 1**).

Publications describe the different types of evidence synthesis methods that have been used in research in health related (6, 7), and agri-food public health areas (14). These studies interestingly do not include CATs as a type of review, which may be an oversight, or indicative of why and how they are used.

ORIGIN OF CRITICALLY APPRAISED TOPICS

The CAT concept was developed by a group of internal medicine fellows at McMaster University, Canada (15) and refined in

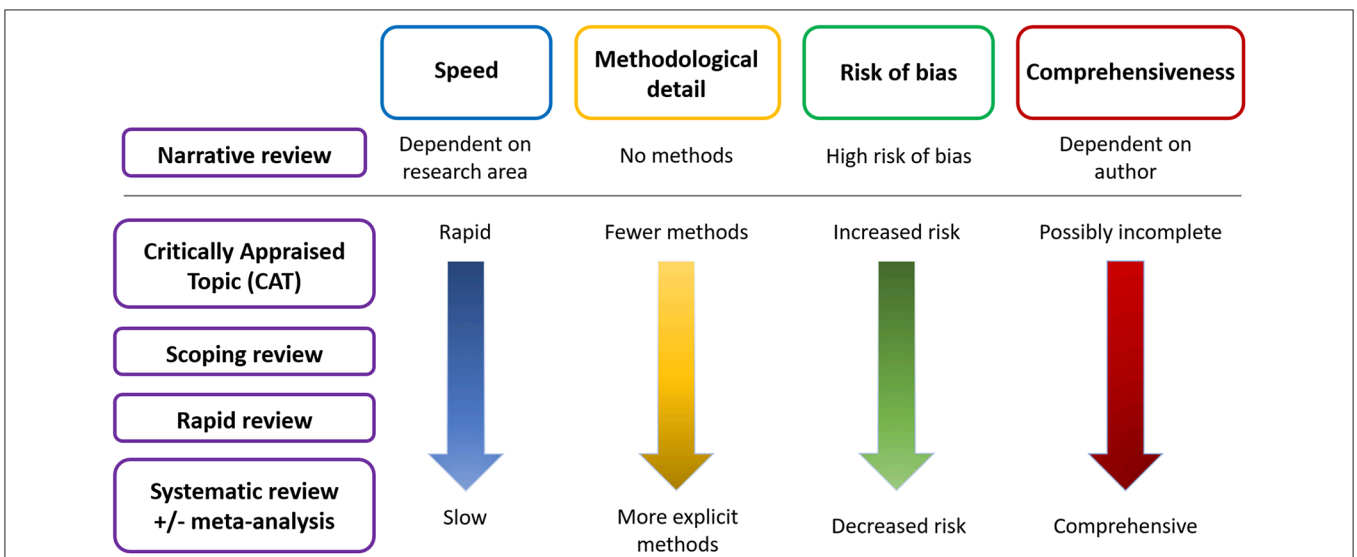


FIGURE 1 | Schematic of the main differences between the types of literature review.

collaboration with a clinical group at Oxford University in the UK (16). CATs were created so fellows could add value to discussions during case rounds and journal clubs (17). It was felt that for busy clinicians, spending a lot of time trying to keep up to date with the wealth of literature was challenging, and traditional methods of searching and reviewing were not applicable (16). Furthermore, for evidence-based medicine (EBM) to be implemented successfully into clinical practice, access to relevant evidence needed to be quickly and easily accessible at the point of patient care (18). CATs helped clinicians learn the skills to search for relevant evidence, critically appraise and write evidence summaries—fundamental skills to practice and teach evidence-based medicine (17). The first CAT process was published in 1993 (19) with the first peer-reviewed article about CATs published in 1995 (17).

The "quick and dirty" applied approach of a CAT makes it versatile and practical to be translated to other disciplines. Physiotherapy, occupational therapy, dermatology, urology, radiology, nursing (9, 11–13, 20), management (21), and education (22) have embraced the CAT approach. The first mention of veterinary professionals using a CAT format was by Cockcroft and Holmes (2); according to the authors, veterinary CATs did not exist at that time. Soon after, a discussion followed about the role of CATs in veterinary education by Hardin and Robertson (23).

USES OF THE METHOD IN VETERINARY MEDICINE

CATs are primarily used in veterinary clinical practice to answer clinical queries resulting from specific cases or conundrums (13). These could be in relation to the case itself, the clinical professionals' knowledge or familiarity with treatments, diagnostic tests, management regimes or surgical approaches, or questions arising from the client. The CAT methodology has been described as a way of closing the gap between clinical research and clinical decision making (15).

CATs are also used in veterinary undergraduate and post-registration education (9, 24) to investigate a clinical question by teaching searching skills, critical appraisal of scientific literature, and the principles of EBVM (23, 25). This is important as research suggests veterinary clinicians (26), mirroring those in other disciplines (9), do not always use an evidence-based approach (e.g. using peer reviewed publications) when finding literature to aid clinical decision making. This is despite EBVM being increasingly recognized as a core skill for all practitioners. The value of the CAT approach in teaching EBVM and critical appraisal skills has been recognized by a variety of veterinary educators globally (3, 27–29).

Other uses in veterinary medicine for CATs are those relevant to any structured review of the literature, including identification of knowledge/research gaps (24, 30), preparation for research grant applications and for informing policy (14, 31).

WHAT ARE THE STEPS IN UNDERTAKING A CRITICALLY APPRAISED TOPIC?

For clinicians, it is useful to think of the CAT process in sequential steps or stages (32, 33). A schematic of the CAT process can be seen in **Figure 2.**

The CAT process is explained in the steps below, using an example to highlight key points, with an overall summary of the example demonstrated in **Figure 3**. Additional information for those more experienced in the CAT methodology is provided in the pop out boxes with double-lined surrounds.

1. Define CAT Question Using Structured Approaches

Transforming a clinical question into a searchable query can be daunting (34). One of the ways to facilitate this process is by using a defined question format. A PICO question (13), where PICO stands for Patient, Intervention, Comparator, Outcome, are the important components a searching strategy should contain (35) if the question relates to treatment efficacy or interventions (for example drugs, vaccines, or surgical procedures). If the question relates to the accuracy of diagnostic tests, then a slightly different format might be appropriate e.g., PIT—Population, Index Test, Target condition or disease (5). Alternative formats for clinical question including prevalence of disease, etiology and comorbidities are described by O'Connor and Sargeant (5).

The PICO format is often illustrated as:

> In [patient group] does [intervention and comparator] result in [outcome]

The following clinical scenario will demonstrate the steps of the CAT process.

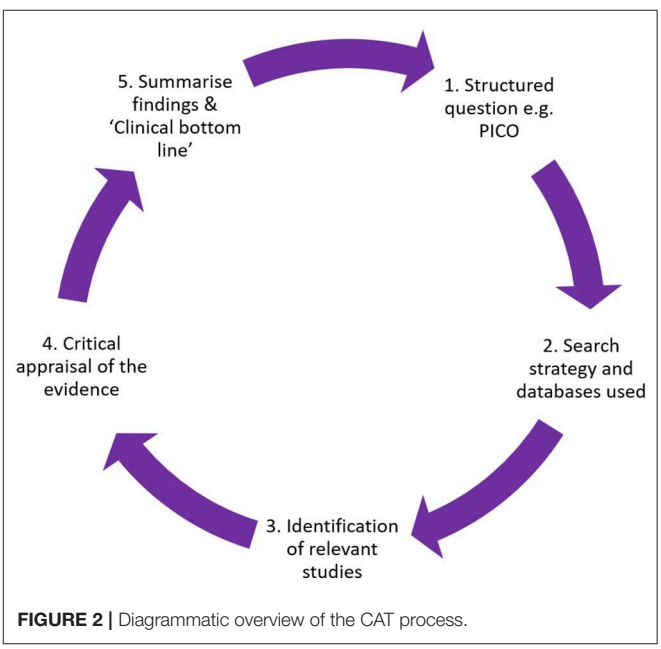


FIGURE 2 | Diagrammatic overview of the CAT process.

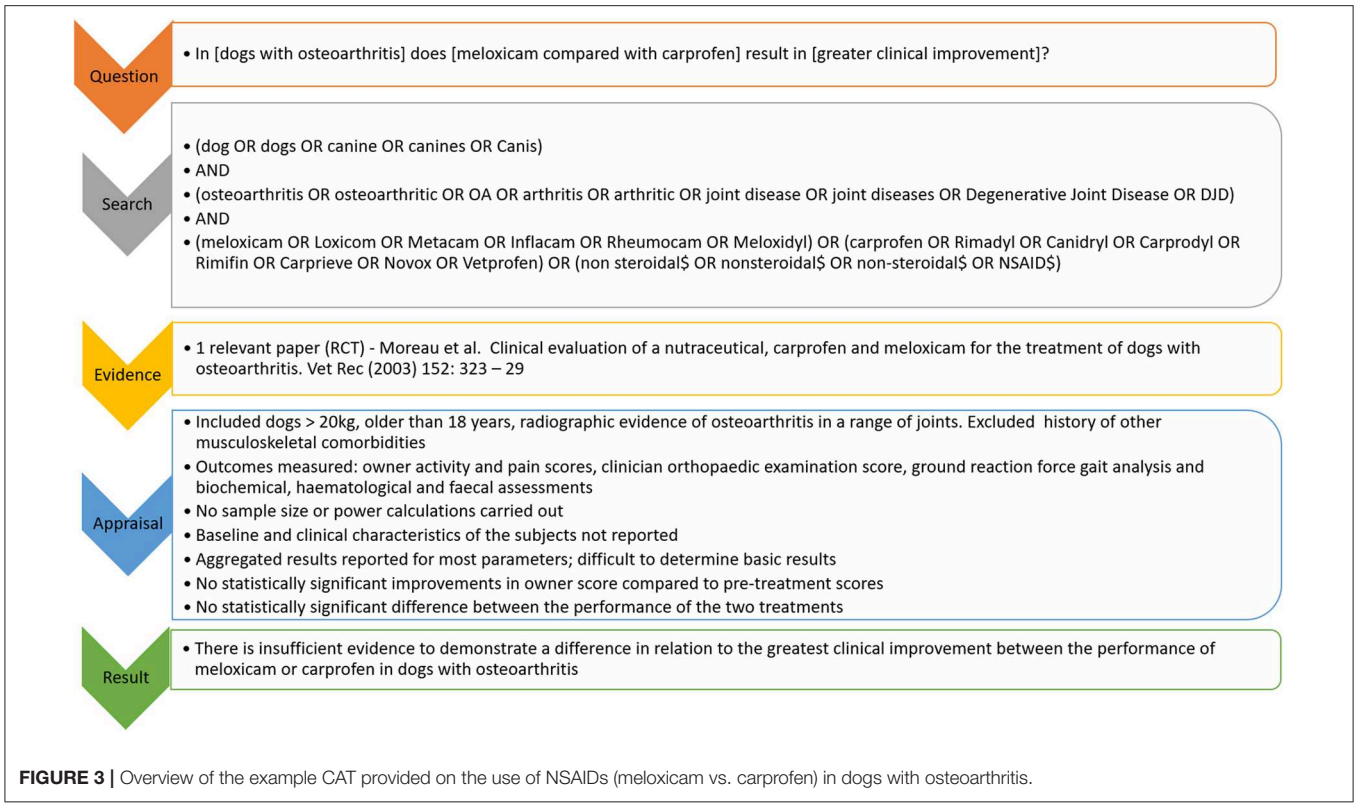


FIGURE 3 | Overview of the example CAT provided on the use of NSAIDs (meloxicam vs. carprofen) in dogs with osteoarthritis.

You have been treating Miley, a 12-year-old Doberman, for osteoarthritis for the past two years. Her owners bring her in for a check-up. On clinical examination you find further reduction in her range of movement, and some signs of pain when you manipulate both of her hind limbs. She is currently on carprofen. Miley's owner asks about meloxicam, as one of the dogs at the park where he walks Miley receives it for a similar problem. You wonder whether Miley may show a greater improvement in clinical signs if she is treated with meloxicam instead of carprofen.

In this clinical scenario, the PICO question might be:

P = Patient group (dogs with osteoarthritis)
I = Intervention (meloxicam)
C = Comparator (carprofen)
O = Outcome (greatest clinical improvement)

In [dogs with osteoarthritis] does [meloxicam compared with carprofen] result in [greater clinical improvement]?

It is possible that further defining the patient group (e.g. forelimb osteoarthritis vs. osteoarthritis) and outcome (e.g. lameness determined by a visual analog scale vs. general clinical improvement) would permit the evidence to be evaluated for applicability more specifically to the clinical case in front of the veterinary professional.

By converting the scenario to a structured PICO format, a search strategy can be focused to answer the question, and appraisal of the evidence (see section below) can focus on the applicability as it relates to the specific question. For further information about searching see the box entitled "General references for defining a question."

General references for defining a question:
De Brun C, Pearce-Smith N. *Searching Skills Toolkit: Finding the Evidence*. Oxford, UK: Wiley-Blackwell. ISBN: 9781118463130 (2009).
EBVM Learning "Ask" module (http://www.ebvmlearning.org/ask/)
EBVM Toolkit 1 (https://knowledge.rcvs.org.uk/document-library/ebvm-toolkit-1-asking-an-answerable-clinical-question/)
PICOvet website (https://pico.vet/index.html)

2. Creating a search strategy

Identifying Search Terms

The PICO question can then be used to search for published evidence relating to the clinical scenario it describes. The first step is identifying search terms that will find the greatest number of relevant publications whilst omitting those that are irrelevant (2). Publications may be inconsistent in the terms used in their titles and abstracts to describe the same thing, so creating a list of synonyms for each PICO component will help to ensure that relevant material is located. By being as comprehensive as possible within each of the P, I, C and O components, the greatest amount of relevant material can be identified.

The search terms identified for the example PICO question are shown in **Table 1**.

In veterinary medicine, the patient group makes up two different sets of terms: the species, and the condition

TABLE 1 | Search terms identified for the PICO question "In [dogs with osteoarthritis] does [meloxicam compared with carprofen] result in [greater clinical improvement]?"

Patient		Intervention	Comparator	Outcome
Species	**Condition**			
Dog	Osteoarthritis	Meloxicam	Carprofen	Clinical
Dogs	Osteoarthritic	Loxicom	Rimadyl	improvement
Canine	OA	Metacam	Canidryl	
Canines	Arthritis	Inflacam	Carprodyl	
Canis	Arthritic	Rheumocam	Rimifin	
	Joint disease	Meloxidyl	Carprieve	
	Joint diseases	Non steroidal	Novox	
	Degenerative	Nonsteroidal	Vetprofen	
	Joint Disease	Non-steroidal	Non steroidal	
	DJD	NSAID	Nonsteroidal	
		NSAIDs	Non-steroidal	
			NSAID	
			NSAIDs	

of interest. The acronym SPICO has been suggested for veterinary medicine (32), starting with "Species" before "Patient group." Note the separate search terms for plurals (e.g. dog, dogs), synonyms (e.g. osteoarthritis, arthritis, degenerative joint disease), and acronyms (e.g. NSAID for non-steroidal anti-inflammatory drug) in the example search terms. Other general considerations are to include "colloquial" terms (e.g. milk fever for hypocalcaemia; cherry eye for nictitans gland prolapse) and eponyms (e.g. Johne's disease for paratuberculosis). When considering synonyms, active ingredients of products (e.g. meloxicam) are the most important terms to look for, although trade names (e.g. Metacam) can also be searched. However, registered trade names differ between countries and, although they may be included as synonyms, they should not be solely relied upon.

Another technique to help with searching inclusivity is truncating or stemming a search term. This is indicated by the addition of a non-letter character, often * or $ depending on the database. Truncated terms can save time when it is likely a number of relevant terms will have the same primary structure (e.g. desex* or desex$ instead of searching for desex, desexed, desexing). These symbols can also be used in the middle of terms to search for different spellings (e.g. "sterili$ation" could be used to represent both the English "sterilisation" and American "sterilization" spellings); this is termed a wild card. Consult the help documentation for each database searched for guidance.

Whilst it is important to identify outcome terms for the PICO as these will assist in determining which of the results are most appropriate, they are often not included in the search. Results from a search of the Patient, Intervention, and Comparison typically yield a sufficiently small number of results that are easily and quickly assessed. Additionally, outcomes may not be clearly defined, it may be difficult to identify all relevant terms for outcomes, and the more concepts that are combined, the greater the risk of excluding a relevant article. Being as specific as possible with the "O" or outcome in the PICO is also useful and important in the appraisal phase of evidence reviews (see section Appraisal of the Evidence).

Structuring the Search and the Use of Boolean Operators

Although the CAT methodology is quite structured, there is a degree of choice and flexibility in how the search is carried out, depending on the timespan available and anticipated amount of evidence. To create a search that is broad ("sensitive") yet relevant ("specific"), terms must be combined in an appropriate way (36). Best practice is to combine search terms and their synonyms using the Boolean operators AND and OR (12, 37); this programs the online search to retrieve relevant results. As a rule, "OR" is used when combining *within* components (e.g. all the patient terms), whilst "AND" is used when combining *separate* components (e.g. patient and intervention term lists) to assure that each component is present in the search results. Capitalizing "OR" and "AND" to denote them as search commands is best practice because it can affect the results returned in some search interfaces.

An additional consideration centres on the differing opinions as to whether the intervention and comparator components should be combined using the Boolean "OR" term. This permits citations to be identified if only one of the two components are mentioned in the abstract. Information specialists, or librarians, have specialist training and are highly skilled in generating searches that optimize the chances of identifying all relevant publications. It is best practice to seek guidance from them whether for training to conduct your own searches, or as collaborators.

The search strategy for the above scenario might appear as follows:

(dog OR dogs OR canine OR canines OR Canis)
AND
(osteoarthritis OR osteoarthritic OR OA OR arthritis OR arthritic OR joint disease OR joint diseases OR Degenerative Joint Disease OR DJD)
AND
(meloxicam OR Loxicom OR Metacam OR Inflacam OR Rheumocam OR Meloxidyl) OR (carprofen OR Rimadyl OR Canidryl OR Carprodyl OR Rimifin OR Carprieve OR Novox OR Vetprofen) OR (non steroidal$ OR nonsteroidal$ OR non-steroidal$ OR NSAID$)

Use of AND allows papers to be identified that contain terms from all components of the search, identifying the most relevant citations, as can be seen in **Figure 4**. In the results, the following must be present: any term from species, any term from patient, any term from intervention, any term from the comparator.

Literature Databases

Once a search strategy has been created, searching can commence within a literature or bibliographic database. These differ from searching the internet using a search engine (e.g. Google or Google Scholar) in two important ways. Bibliographic databases contain journal articles that are not generally available online or accessible via internet search engines. Coverage by internet search engines is not transparent and changes frequently.

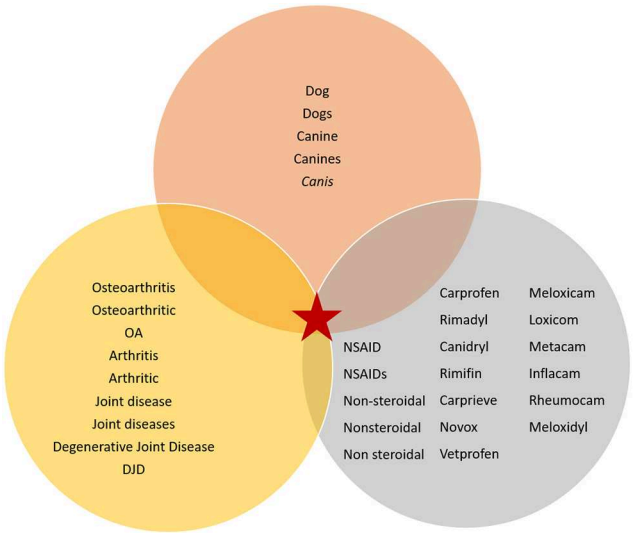


FIGURE 4 | Venn diagram illustrating the interaction between the different search terms within the search components for the example scenario (denoted by a red star).

A number of bibliographic databases exist. Research suggests at least two databases should be searched, including CAB Abstracts since it contains the most comprehensive database for veterinary topics (38). The data in CAB Abstracts are available for subscription by institutions (39), and for individual subscription as VetMed Resource (40) which includes bibliographic records, limited full-text, and links to free and subscription articles. PubMed, from the US National Library of Medicine, is a freely available bibliographic database that covers biomedical sciences including a core of veterinary medicine information (41). It includes the MEDLINE database and additional bibliographic records and links to free and subscription articles.

For those employed at a university or corporation, check with your information specialists or librarians to find the databases available to you. For those not affiliated with an institution, collaboration with individuals at universities or obtaining practice or individual subscriptions to databases [e.g. VetMed Resource (40)] is useful. Some professional bodies offer access to databases as a member benefit. It would be pertinent for veterinary professionals to consider membership to relevant initiatives such as the RCVS Knowledge Library (42) which offers training in literature searching. Other cost-effective options are available (37).

Searching Using Database Specific Subject Headings

Search results may be improved by the inclusion of standardized terms in the search (43). These terms are specific to each bibliographic database. The content of each publication being indexed is identified, assessed, and assigned a standardized database specific term, often called a subject heading. These are "umbrella" terms for a given concept and are organized in a database-specific thesaurus. PubMed and MEDLINE use MeSH (Medical Subject Headings), CAB Abstracts uses the CABI Thesaurus. Consult the help documentation provided by the databases for guidance. Not all CAT guidance recommends the use of subject heading searches (11, 12) but when used, it is likely to improve the sensitivity of searching strategies (43) and therefore should be carried out if possible. For further information about subject headings, see the box entitled "General references about using subject headings."

General references about using subject headings

EBVM Toolkit 2 (https://knowledge.rcvs.org.uk/document-library/ebvm-toolkit-2-finding-the-best-available-evidence/)

EBVM Learning Acquire (http://www.ebvmlearning.org/acquire/)

PubMed for Veterinarians (https://www.tamucet.org/product/pubmed-for-veterinarians/)

3. Identification of Relevant Studies

After the search has been carried out, the next step is to identify publications that can be used to answer the CAT question. Firstly, the citation must be relevant to the PICO question—it must contain all the components, including the outcome of interest. This assessment begins by looking at the title of each citation. If the title does not sound relevant, the citation is excluded and the next one is assessed (37). If the title is potentially relevant, the abstract is assessed for further detail. If the abstract is relevant, the full text of the article is scanned. If the full text article is not available, the citation may be excluded or further work undertaken to obtain a complete copy (37). There is a flow diagram that appears in White and Larson (44) that can help to facilitate the process described above.

Secondly, exclusions might apply to ensure the citations are as evidence-based as possible. For example, those citations that are not peer-reviewed (e.g. conference proceedings, textbooks, theses), do not contain evidence of research methodology (e.g. narrative reviews), or are carried out in a non-applied setting (e.g. *in vitro* research) may be excluded (36). Often if the full text version of a paper is in a language in which the authors are not sufficiently fluent, it is excluded due to the lack of time for translations in the rapid CAT process.

It is possible that at the end of this stage, no relevant peer-reviewed citations are found, or the material found provides insufficient confidence that the findings are valid. The searching strategy could be amended (e.g. using "OR" between the I and C components instead of "AND") to "widen the net." If this is not successful, the process of a traditional CAT ends here. Some published CATs include searches that don't return any citations to demonstrate evidence gaps (45). If the search retrieves no results but clinical decisions need to be made about a case, other forms of evidence such as conference proceedings, textbooks, narrative reviews and expert opinion could be used instead (46). Publications looking at the PICO topic as it relates to other species (including humans), or those containing *in vitro* studies could be investigated.

In the example scenario above, a MEDLINE search returned 345 citations, one of which was relevant. No papers were excluded because they were not in English, 11 papers were excluded as they were narrative reviews, conference proceedings or related to *in vitro* research, and 333 were excluded because they did not meet all components of the PICO question. A CAB Abstracts search

returned 412 citations, one of which was relevant (the same paper as in the MEDLINE search). One paper was excluded as it was not in English, nine as they were narrative reviews, conference proceedings or related to *in vitro* research, and 401 were excluded because they did not meet all components of the PICO question. This left a total of one relevant paper from the two database searches, Moreau et al. (47).

4. Appraisal of the Evidence

One of the most important parts of the CAT process is the appraisal of the evidence. This assesses the study design and its execution (48). Often there is an assumption by medical and veterinary professionals that if something is published in a scientific journal, it is automatically valid and high quality. However, the publishing and peer review process is not flawless (49–51) and not all published articles are of equal quality (52). Therefore, it is important that all publications undergo an assessment of how they were conducted.

The first step in this process is to identify the study design and assess its place in the evidence "hierarchy" (24). All study designs have a degree of bias associated with them, but some are considered more objective than others (8). A number of schematics rank the study designs according to their inherent level of bias (hence "hierarchy") in a "pyramid" [(2); **Figure 5**] or "staircase" of evidence (52). Study designs at the top of the pyramid are theoretically the least biased (e.g. systematic reviews and meta-analyses), with bias increasing toward the bottom (e.g. personal anecdotes) (24). The pyramid shape also indicates that the majority of evidence sources are at the bottom, with fewer, less biased studies at the tip of the pyramid (54).

However, the pyramid can be followed too strictly, ignoring the point that the "ideal" study design to answer a specific question relates to the type of question that is being asked.

> Concerns raised include whether a case-control design is of lesser evidentiary value than cohort studies and whether the terms cohort, case control, and case series can be used to "filter" out studies of lower evidentiary value (55). Additionally, this common version of the pyramid is geared toward questions of treatment comparisons (SRs, meta-analyses and randomized controlled trials appearing at the top of the pyramid). Where a question relates to other types of clinical question, such as establishing disease prevalence, the "hierarchy" here no longer applies—for example, cross sectional studies are more appropriate to conduct in this case than randomized controlled trials (56). Additionally, there are an increasing number of qualitative research studies being undertaken in veterinary medicine; it is difficult to know where to integrate these studies into the traditional pyramid hierarchy.

It can be difficult to determine what type of study design has been carried out; the stated study design may not be correct (57, 58), which can leave CAT authors uncertain as to how to approach reading the paper. There are a number of resources that contain a good description of common study designs (1, 8, 59), including some with flow diagrams for helping to determine what type of study design has been used (8, 60, 61).

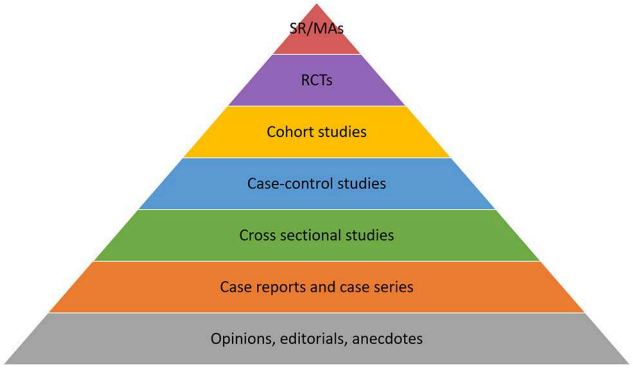


FIGURE 5 | Pyramid of evidence, modified from Phillips (53). SR, Systematic reviews; MA, Meta-analysis; RCTs, Randomised controlled trials.

The second step in the process is to determine whether the study design has been executed in the appropriate manner; this assessment is termed "critical appraisal" (48). This is often undertaken using structured worksheets which contain questions tailored to the specific study design (13). There are many different resources that can be used for this process, but all are fundamentally similar in the questions they address. Some examples from the medical and veterinary field are highlighted in the box entitled "General references for appraising the evidence."

General references for appraising the evidence

Veterinary—

CEVM website (https://www.nottingham.ac.uk/cevm/evidence-synthesis/resources.aspx)

EBVM toolkit, RCVS Knowledge (https://knowledge.rcvs.org.uk/evidence-based-veterinary-medicine/ebvm-toolkit/)

Dean RS. How to read a paper and appraise the evidence. *Practice.* (2013) 35:282–5. doi: 10.1136/inp.f1760

Downes MJ, Brennan ML, Williams HC, Dean RS. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ Open.* (2016) 6:e011458. doi: 10.1136/bmjopen-2016-011458

Moberly HK. How to read and appraise veterinary articles. *Texas Vet.* (2019) 81:54. uri: 1969.1/178285

Pinchbeck GL, Archer DC. How to critically appraise a paper. *Equine Vet Educ.* (2020) 32:104–9. doi: 10.1111/eve.12896

Medicine—

Centre for Evidence Based Medicine (https://www.cebm.net/2014/06/critical-appraisal/)

CASP (https://casp-uk.net/casp-tools-checklists/)

Joanna Briggs Institute (https://joannabriggs.org/ebp/critical_appraisal_tools)

How to read a paper series, British Medical Journal (https://www.bmj.com/about-bmj/resources-readers/publications/how-read-paper)

Crombie IK. *The Pocket Guide to Critical Appraisal*. London: BMJ Publishing Group (2009).

Greenhalgh T. *How to Read a Paper: the Basics of Evidence-Based Medicine*. 5th ed. Chichester, UK: Wiley-Blackwell (2014).

While specific questions need to be answered based on the study's design, there are key, easy questions that should be asked of all study types. These are (62):

- Does this study address a clearly focused question?
- Did the study use valid methods to address this question?
- Are the valid results of this study important?

- Are these valid, important results applicable to my patient or population?

The Centre for Evidence-Based Medicine takes a time-efficient approach to the answers to these questions, saying that if the answer is no to any of them, clinicians should avoid reading the rest of the paper as it is not relevant (62).

Veterinary professionals can worry that appraisal will be too difficult and may need advanced understanding of statistics. In reality, critical appraisal relies on the application of common sense in conjunction with an appraisal template with much of the focus on the study design, not the statistics. For example, of the 27 questions posed in the randomised controlled trial (RCT) critical appraisal sheet developed by the Centre for Evidence-based Veterinary Medicine (CEVM), only four relate to statistical calculations (63). None require the appraiser to carry out any statistical tests. There are a number of easy to understand statistical reference guides that could assist professionals to interpret common types of analyses (64–66) that could be used alongside the structured worksheets to assist them in interpreting the study results. Alternatively, assistance or further training could be sought (2, 67), but this level of higher knowledge is rarely required. Therefore, veterinary professionals should be able to appraise the vast majority of important features within each study in order to draw meaningful conclusions. If the study is well-conducted and well-reported, it should be easy to critically appraise.

In the given scenario, the paper that was identified (47) was a randomised controlled trial. After appraisal using the RCT critical appraisal sheet from the CEVM (63), the main points to note in relation to the study were:

- Prior to the study commencing:
  - There was no assessment of how many animals would be required prior to the study commencing (e.g. no sample size or power calculations were presented).
- Once the study had commenced:
  - The study focused on dogs weighing more than 20 kg and were older than 18 months of age with radiographic evidence of osteoarthritis in a range of joints. Subjects were excluded if there was history of other types of musculoskeletal comorbidities.
  - Outcomes measured were owner activity and pain scores, clinician orthopedic examination score, ground reaction force gait analysis and biochemical, haematological and faecal assessments.
  - Baseline characteristics and clinical characteristics of the subjects were not reported.
  - Aggregated results were reported for most but not all parameters; it was difficult to determine basic results as a consequence.
  - There were no statistically significant improvements in owner score compared to pre-treatment scores. The exception was a subset of dogs with stifle disease in the Metacam group ($n = 6$) who showed an improvement at day 30 only (not at day 60). There was no statistically significant difference between the two treatments for this measure.
  - Within each treatment group, there were statistically significant improvements in clinician score (at day 30 only), and in selected ground reaction force measures compared to pre-treatment scores. There was no statistically significant difference between the performance of the two treatments.

5. Summarise Findings and "Clinical Bottom Line"

The last part of the process is an overall assessment of all the evidence appraised. There is no standard way of amalgamating results from appraisals in the CAT format (36) but it becomes easier with practice. Challenges include comparisons of different types of study design (e.g. a randomized controlled trial and cohort study), and where different studies report conflicting answers to the question. Conflicting answers could be related to the varying abilities/characteristics and biases inherent in the different study types (35), or because different populations have been studied (e.g. shelter animals vs. owned animals; a study based in Australia vs. a study based in Canada). This is where the judgement of the veterinary professional becomes important. It is a common occurrence that, even after reading the evidence, there is no clear, definitive conclusion. It should be noted that such an outcome is distinct from a conclusion stating that there is no effect of the intervention.

In the given scenario, the study weaknesses were felt to be substantial enough to conclude it was not possible to answer the clinical question. The clinical bottom line was that there was insufficient evidence to demonstrate a difference in relation to the greatest clinical improvement between the performance of meloxicam or carprofen in dogs with osteoarthritis. For an overall summary of the example CAT provided here, refer back to **Figure 3**.

PUBLISHING A CAT

Production of the CAT can be carried out by more than one author (45) to increase objectivity and reduce bias. Once the question and search strategy are agreed, multiple authors may independently search the literature and/or, more commonly, agree on any relevant studies. They reach a consensus on which studies to include and then independently appraise their quality, before collaborating again to summarize the findings and arrive at the clinical bottom line. There is a lack of guidance on reporting CATs in the literature, and those that do exist tend to be journal specific. At the time of writing, the Veterinary Evidence journal had the most comprehensive guidance for reporting Knowledge Summaries (a form of CAT) (33), with minimal guidance provided by Equine Veterinary Education (68). It is recommended also to look at the examples following this section for further

guidance on reporting. Journals such as the Veterinary Record (publish BestBETs—a form of CAT—and other formats), BMC Veterinary Research, Equine Veterinary Education and Veterinary Evidence (publish Knowledge Summaries) have published CATs previously.

Good Examples of Critically Appraised Topics From Medicine and Veterinary Medicine

There are a number of excellent examples of CATs and resources available to help facilitate the construction of CATs, both in the medical and veterinary fields. This section will focus on published examples of CATs, collections of existing CATs, and website resources that can be utilized to construct CATs. The applied nature of CATs means that many of the most useful "how to" resources are not published in peer-reviewed journals, but on university webpages, open access online tutorials or online databases.

Medicine

Over time there have been a number of medical CAT databases in existence; in 2005 there were at least 13 different places where medical CATs appeared (69); it is unknown how many of these are still regularly contributed to. Software was developed to be able to search simultaneously across a number of different CAT databases ["CAT crawler"; (70)], but widespread use of this is not evident in the literature.

A good example of a working database of CATs is BestBETs (www.bestbets.org). This database was constructed by emergency clinicians working at the Manchester Royal Infirmary in the UK, in response to a lack of high quality evidence for some of what was seen regularly in emergency care (71), hence the use of the term "Best Evidence Topics (BETs)." Some of these BETs are also published in peer-reviewed journals. The topics covered in this database have expanded to include other specialties besides emergency medicine, including cardiothoracics and paediatrics.

Veterinary Medicine

There are numerous different formats of CATs available in veterinary medicine, most of which have emerged over the past 10 years. There are some differences between these formats in relation to how the review question has come about, what format the review is available in (e.g. on a website, published literature), and how the "review" component of each format occurs (e.g. number of authors, reviewers etc.), but they essentially follow the same process. The advantage for veterinary professionals is that there are several CAT collections available to utilise for decision making in clinical practice. The collections of veterinary CATs available at the time of article preparation are listed alphabetically in **Table 2**. The majority of these are freely available, although not all appear to be current and are being updated at variable frequencies. Published examples of CATs and useful web sources to help create CATs can be seen in the inset boxes below.

Published examples of veterinary CATs:

There are several good examples of veterinary CATs that have been published in the literature. Two can be seen here, both of which are free to view. These examples demonstrate a contrast in relation to the types of question and approaches that can be used under a CAT format.

Finka LR, Ellis SLH, Stavisky J. A critically appraised topic (CAT) to compare the effects of single and multi-cat housing on physiological and behavioral measures of stress in domestic cats in confined environments. *BMC Vet Res.* (2014) 10:73. doi: 10.1186/1746-6148-10-73

This CAT contributed to the development of welfare guidelines for unowned cats (72).

Olivry T, Mueller RS, Prelaud P. Critically appraised topic on adverse food reactions of companion animals (1): duration of elimination diets. *BMC Vet Res.* (2015) 11:3. doi: 10.1186/s12917-015-0541-3

Useful web sources:

Medicine:

"How to" resources—

Centre for Evidence Based Medicine CATmaker: (https://www.cebm.net/2014/06/catmaker-ebm-calculators/)

Physiopedia: (https://www.physio-pedia.com/Critically_Appraised_Topics)

Veterinary medicine:

Other CATs—

Healthy Feet website: (https://www.cattle-lameness.org.uk/critically-appraised-topics/)

BMC adverse food reaction CATs: (https://www.biomedcentral.com/collections/catsfoodreactions)

SOME OF THE LIMITATIONS AND MISINTERPRETATIONS ASSOCIATED WITH CATs

The main limitation associated with the use of the CAT methodology is the "quick and dirty" nature of the process. Due to the rapid approach, the process is not as detailed nor in depth as other types of review, and therefore there is more potential to miss relevant evidence sources (9). This may mean that a CAT may not be representative of the totality of the evidence in existence on a particular topic (36). In addition, the questions asked when using a CAT format are usually narrow and tend to be very specific to a clinical scenario or experience (9). This sometimes limits the ability to translate findings to a wide variety of situations. However, there are so few evidence-based resources and limited funding for CAT resources for veterinary professionals that any structured reviews can be of benefit to clinical decision making. As with any type of review publication, they can become outdated and should be re-assessed regularly (9).

There are other pitfalls associated with this methodology which are inherently related to structured reviews generally. If only "colloquial" terms are used (those used locally or regionally) to describe diseases/conditions/procedures then it is more likely a CAT author from a different part of the world may miss a relevant publication. For example, the term "tup" can be used to describe a male sheep in the UK; in other countries this term is not generally used. The majority of known CAT collections in veterinary medicine are published in English, and to the authors'

TABLE 2 | Collections of veterinary CATs.

Name of CAT collection	References	Start date of CATs	Date of last CAT	Approach	Type of source	Frequency of updates
Banfield Applied Research and Knowledge (BARK) CATs website	https://www.banfield.com/veterinary-professionals/resources/research/cats	Nov 2009	2013	Single author reviews	Free to view	Unknown
BestBETs for Vets website (and a selection of these in the Veterinary Record journal)	http://bestbetsforvets.org/	Sept 2013	Current at time CAT review article published	Multi-author reviews	Free to view	Every 2 years
Equine Veterinary Education journal	https://beva.onlinelibrary.wiley.com/hub/journal/20423292/homepage/critically_appraised_topics_for_clinical_evidence_in_equine_practice.html	March 2015 (for clinical evidence series of CATs)	2020	Single author reviews	Free to view	No set timeline
Veterinary Evidence journal (called Knowledge Summaries)	https://veterinaryevidence.org/index.php/ve/index	Oct 2015	Current at time CAT review article published	Single author reviews	Free to view	Most popular automatically updated every 2 years; others when required
Veterinary Prescriber website	https://www.veterinaryprescriber.org/	Mar 2014	Current at time CAT review article published	Multi-author reviews	Subscription based	Variable; 3 years or more
"Where's the evidence" series in the Journal of the American Veterinary Medical Association in conjunction with the Evidence Based Veterinary Medicine Association	https://avmajournals.avma.org/loi/javma/	Nov 2009	Aug 2011	Multi-author reviews	Subscription based	Unknown

knowledge, none of the reviews in these databases go to the extent of searching for non-English publications for inclusion. This is a distinct limitation (73) but is also likely to be related to the rapid nature of these reviews in relation to the delay it may take for additional searching and translation to be undertaken.

For relevant studies to be identified, published research must be indexed correctly. Information specialists rely on authors identifying the most appropriate key words for their publication and ensuring the most important terms are included in the title and abstract. It also depends on the terminology used to describe disease conditions or procedures. Additionally, depending on the database in question, some of the indexing of veterinary related publications is done by personnel who may not necessarily be familiar with some of the conditions that afflict animals. Automated indexing systems can both omit relevant subject headings from a record which can impact on retrieval or include erroneous subject headings. All of the above can impact on whether specific publications are returned after a structured search has been performed.

There are sometimes misconceptions by veterinary professionals in relation to these clinically relevant reviews of the literature, analogous to those held by some medical professionals in relation to clinical guidelines (74). They can be seen as the definitive answer, from which health professionals are only allowed to deviate for good reasons. This can be comforting, particularly to those inexperienced in clinical decision making, such as new graduates. Alternatively, CATs can be regarded as over-prescriptive, too restrictive in scope and even draconian. However, these reviews should always be applied contextually to the patient in front of the decision-maker. If the study populations are substantially different to their own patients, the veterinary professional may deem the CAT irrelevant and choose to ignore it (24). The evidence must be applied within the context of the circumstances of the patient and owner in order for the clinical plan and treatment to have the greatest chance of success.

FURTHER DEVELOPMENTS

Excluding for educational purposes, the role of the CAT appears to have been superseded by SRs, which are often used as the basis for clinical guidelines for medical practitioners [e.g. in the UK, Clinical Knowledge Summaries; (75)]. SRs are a more thorough representation of the existing evidence than CATs, and include both published and, often, unpublished sources of information. For areas not covered by such guidelines (e.g. common questions still to be answered, and areas where it is inherently difficult to undertake unbiased types of study such as randomised controlled trials, for example in emergency medicine), CATs will continue to play an important role in clinical practice. In veterinary medicine, there are unlikely to be large numbers of systematic reviews generated in order to develop clinical guidelines, primarily due to a lack of both suitable research funding and appropriate skills within the veterinary profession. However, many veterinary and nursing undergraduate courses and further education courses for technicians globally include elements of EBVM training (such as how to carry out a CAT) within them (27, 28) and there are also opportunities now for post-registration training and

continuing professional development in these skills (32, 67, 76, 77). This suggests that the skill base may well increase in the future. With some additional work, the CATs undertaken by students that are currently kept internally within institutions could become publicly available CAT collections. Alternatively, with some assistance from educators, student CATs could be published in veterinary journals. There are awards available for students to publish CATs currently (78), which should facilitate this process.

With the creation of more CAT collections in the veterinary sphere (**Table 2**), professionals can use CATs without requiring the same skills needed to generate them. To facilitate carrying out the CAT process in clinical practice, adequate time must be given to professionals to be able to perform searches and interpret evidence during their working day. This is a bigger challenge for the profession that must be prioritized moving forwards. With the rise of corporate practice groups, there have emerged roles with the responsibility of ensuring EBVM-based practice. This may accelerate the prioritization of evidence reviewing as part of a veterinary professionals' role, which could increase the demand for CATs and thereby facilitate formation of a more centralized source for professionals.

For busy practitioners, having numerous different CAT collections to search across is suboptimal. In the future it may be that provision of software, such as the "CAT crawler" (70) would overcome this barrier. However, this requires funding for development and maintenance which, for these sorts of resources, is unlikely to be prioritized by funding bodies. Additionally, in order to increase the translatability of the CATs in these collections, adding a patient perspective section may add a different dimension. The CAT undertaken by Wootton et al. (79) that appeared in the British Journal of Dermatology includes such a section, so a template is already in existence that could be utilized. A similarly motivated patient perspective column has recently been initiated as a feature in the Veterinary Record journal (80) which demonstrates the power of the client's voice.

CONCLUSION

The CAT framework is still a current and useful process for veterinary professionals to use primarily for evidence-based clinical decision making and for undergraduate and post-registration training. With the provision of new CAT collections that can be utilized often at no cost, there are good options available for those in clinical practice who do not yet have the skills to generate CATs themselves. All veterinary professionals, with regular practice, have the ability to successfully navigate the CAT process. However, time must be given to those in clinical practice for the development of these skills so that more CATs can be generated, facilitating excellent evidence-based care of clients and their animals.

AUTHOR CONTRIBUTIONS

MB, LC, HD, and LM were involved in creating the framework for the manuscript. MB, SA, ZB, LB, LC, HD, VF, DG, HM, LM, JS, and CW contributed to acquisition of data (publications) for the work. MB wrote the draft manuscript. MB, SA, ZB, LB, LC, HD, VF, DG, HM, LM, JS, and CW contributed to editing the manuscript, read and approved the final manuscript.

ACKNOWLEDGMENTS

Many thanks to the veterinary professionals, students and information specialists we work with for their feedback and support in our endeavors.

REFERENCES

1. Dean RS. How to read a paper and appraise the evidence. *In Pract*. (2013) 35:282–5. doi: 10.1136/inp.f1760
2. Cockcroft PD, Holmes MA. *Handbook of Evidence Based Veterinary Medicine*. Oxford, UK: Blackwell Publishing (2003). doi: 10.1002/9780470690833
3. Arlt SP, Haimerl P, Heuwieser W. Training evidence-based veterinary medicine by collaborative development of critically appraised topics. *J Vet Med Educ*. (2012) 39:111 −8. doi: 10.3138/jvme.1111.112R
4. Kelly AM, Cronin P. How to perform a critically appraised topic: part 2, appraise, evaluate, generate and recommend. *AJR Am J Roentgenol*. (2011) 197:1048–55. doi: 10.2214/ajr.11.7249
5. O'Connor A, Sargeant J. Research synthesis in veterinary science: narrative reviews, systematic reviews and meta-analysis. *Vet J*. (2015) 206:261–7. doi: 10.1016/j.tvjl.2015.8.025
6. Sutton A, Clowes M, Preston L, Booth A. Meeting the review family: exploring review types and associated information retrieval requirements. *Health Info Libr J*. (2019) 36:202–22. doi: 10.1111/hir.12276
7. Grant, MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J*. (2009) 26:91–108. doi: 10.1111/j.1471-1842.2009.00848.x
8. Larson RL, White BJ. Importance of the role of the scientific literature in clinical decision making. *J Am Vet Med Assoc*. (2015) 247:58–64. doi: 10.2460/javma.247.1.58
9. Foster N, Barlas P, Chesterton L, Wong J. 'Critically appraised topics (CATs): One method of facilitating evidence-based practice in physiotherapy'. *Physiotherapy*. (2001) 87:179–90. doi: 10.1016/S0031-9406(05)60604-9
10. O'Connor A, Fajt V. Evaluating treatment options for common bovine diseases using published data and clinical experience. *Vet Clin North Am Food Anim Pract*. (2015) 31:1–15.doi: 10.1016/j.cvfa.2014.11.001
11. Beckers GMA, Herbst K, Kaefer M, Harper L, Castagnetti M, Bagli D, et al. Evidence based medicine IV; how to find an evidence-based answer to a clinical question? make a critically appraised topic! *J Pediatr Urol*. (2019) 15:409–11. doi: 10.1016/j.jpurol.2019.05.009
12. Kelly AM, Cronin P. How to perform a Critically Appraised Topic: Part 1, ask, search, and apply. *AJR Am J Roentgenol*. (2011) 197:1039–47. doi: 10.2214/ajr.09.7205
13. White S, Raghavendra P, McAllister S. Letting the CAT out of the bag: contribution of Critically Appraised Topics to evidence-based practice. *Evid Based Commun Assess Interv*. (2017) 11:27–37. doi: 10.1080/17489539.2017.1333683
14. Young I, Waddell L, Sanchez J, Wilhelm B, McEwan SA, Rajic A. The application of knowledge synthesis methods in agri-food public health: recent advancements, challenges and opportunities. *Prev Vet Med*. (2014) 113:339–55. doi: 10.1016/j.prevetmed.2013.11.009
15. Wyer PC. The critically appraised topic: closing the evidence-transfer gap. *Ann Emerg Med*. (1997) 30:639–40. doi: 10.1016/S0196-0644(97)70083-0

16. Sequeira P. Critically appraised topics (CATs): bringing evidence to the chairside (1). *Evid Based Dent*. (2002) 2:107–9. doi: 10.1038/sj.ebd.6400075
17. Sauve S, Lee HN, Meade MO, Lang JD, Farkouh M, Cook DJ et al. The critically appraised topic: a practical approach to learning critical appraisal. *Ann Roy Soc Phys Surg Can*. (1995) 28:396–8.
18. Lee HM, Sauve JS, Farkouh ME, Sackett DL. The critically appraised topic: a standardized aid for the presentation and storage of evidence-based medicine. *Clin Res*. (1993) 41:543A.
19. Sackett DL, Straus SE. Finding and applying evidence during clinical rounds: the evidence cart. *JAMA*. (1998) 280:1336–8. doi: 10.1001/jama.280.15.1336
20. Callendar J, Anstey AV, Ingram JR, Limpens J, Flohr C, Spuls PI. How to write a critically appraised topic: evidence to underpin routine clinical practice. *Br J Dermatol*. (2017) 177:1007–13. doi: 10.1111/bjd.15873
21. CEBMa. *CEBMa Guideline for Critically Appraised Topics*. Amsterdam: Centre-for Evidence-based Management (2017). Available online at: https://www.cebma.org/wp-content/uploads/CEBMa-CAT-Guidelines.pdf (accessed May 27, 2020).
22. Davies P. What is evidence-based education? *Br J Educ Stud*. (1999) 47:108–21.
23. Hardin LE, Robertson S. Learning evidence-based veterinary medicine through development of a critically appraised topic. *J Vet Med Educ*. (2006) 33:474–8. doi: 10.3138/jvme.33.3.474
24. Holmes MA. Philosophical foundations of evidence-based medicine for veterinary clinicians. *J Am Vet Med Assoc*. (2009) 235:1035–9. doi: 10.2460/javma.235.9.1035
25. Steele M, Crabb NP, Moore LJ, Reyher KK, Baillie S, Eisler MC. Online tools for teaching evidence-based veterinary medicine. *J Vet Med Educ*. (2013) 40:272–7. doi: 10.3138/jvme
26. Vandeweerd JEF, Vandeweerd S, Gustin C, Keesemaecker G, Cambier C, Clegg P, et al. Understanding veterinary practitioners' decision-making process: implications for veterinary medical education. *J Vet Med Educ*. (2012) 39:142–51. doi: 10.3138/jvme.0911.098R1
27. Dean RS, Brennan M, Baillie S, Brearley J, Cripps P, Eisler MC, et al. The challenge of teaching undergraduates evidence-based veterinary medicine. *Vet Rec*. (2017) 181:298–9. doi: 10.1136/vr.j3441
28. Janicke H, Johnson MA, Baillie S, Warman S, Stone D, Paparo S, et al. *Creating the Next Generation of Evidence-based Veterinary Practitioners and Researchers: What Are the Options for Globally Diverse Veterinary Curricula? Report for Council for International Veterinary Medical Education*. Washington, QC: Association of American Veterinary Medical College (2019). Available online at: https://www.aavmc.org/assets/site_18/files/civme/civme.toolbox.research%20skills.and.ebvm.pdf (accessed May 27, 2020).
29. Fajt VR, Brown D, Scott MM. Practicing the skills of evidence-based veterinary medicine through case-based pharmacology rounds. *J Vet Med Educ*. (2009) 36:186–95. doi: 10.3138/jvme.36.2.186
30. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. (2005) 8:19–32. doi: 10.1080/1364557032000119616
31. ECDC. *Evidence-Based Methodologies for Public Health*. Solna: European Centre for Disease Prevention and Control (2011). Available online at: http://ecdc.europa.eu/en/publications/Publications/1109_TER_evidence_based_methods_for_public_health.pdf (accessed May 27, 2020).
32. EBVM Learning. *EBVM Learning Tutorial*. London: RCVS Knowledge (2015). Available online at: http://www.ebvmlearning.org/ (accessed May 27, 2020).
33. Veterinary Evidence. *How to Write a Knowledge Summary*. London: RCVS Knowledge (2020). Available online at: https://www.veterinaryevidence.org/index.php/ve/how-to-write-a-knowledge-summary (accessed May 27, 2020).
34. Robertson SR. Refining the clinical question: the first step in evidence-based veterinary medicine. *Vet Clin North Am Small Anim Pract*. (2007) 37:419–31. doi: 10.1016/j.cvsm.2007.01.002
35. Larson RL, White BJ. First steps to efficient use of the scientific literature in veterinary practice. *J Am Vet Med Assoc*. (2015) 247:254–8. doi: 10.2460/javma.247.3.254
36. Dean R, Mackway-Jones K, Wareham K, Grindlay D, Brennan M. BestBETs for vets: a way to improve the odds of delivering high-quality care. *Vet Rec*. (2015) 176:354–6. doi: 10.1136/vr.h1593
37. EBVM Learning. *Acquire*. London: RCVS Knowledge (2015). Available online at: http://www.ebvmlearning.org/acquire/ (accessed May 27, 2020).
38. Grindlay DJC, Brennan ML, Dean RS. Searching the veterinary literature: a comparison of the coverage of veterinary journals by nine bibliographic databases. *J Vet Med Educ*. (2012) 39:404–12. doi: 10.3138/jvme.1111.109R
39. CABI. *Publishing Products*. Wallingford: CABI (2019). Available online at: https://www.cabi.org/publishing-products/cab-abstracts/. (accessed May 27, 2020).
40. CABI. *Vet Med Resource*. Wallingford: CABI (2020). Available online at: https://www.cabi.org/vetmedresource/ (accessed May 27, 2020).
41. *PubMed*. Bethesda, MD: US National Library of Medicine National Institutes of Health (2020). Available online at: https://www.ncbi.nlm.nih.gov/pubmed/ (accessed May 27, 2020).
42. RCVS Knowledge. *Library and Information Services*. London: RCVS Knowledge (2020). Available online at: https://knowledge.rcvs.org.uk/library-and-information-services/ (accessed May 27, 2020).
43. Baumann N. How to use the medical subject headings (MeSH). *Int J Clin Pract*. (2016) 70:171–4. doi: 10.1111/ijcp.12767
44. White BJ, Larson RL. Systematic evaluation of scientific research for clinical relevance and control of bias to improve clinical decision making. *J Am Vet Med Assoc*. (2015) 247:496–500. doi: 10.2460/javma.247.5.496
45. BestBETs for Vets. *BestBets for Vets*. Nottingham: Centre for Evidence-based veterinary Medicine (2020). Available online at: http://bestbetsforvets.org/ (accessed May 27, 2020).
46. Reyher, KK, Brennan, ML. Evidence-based veterinary medicine in action in farm animal practice. *Cattle Prac*. (2015) 23:157–61.
47. Moreau M, Dupuis J, Bonneau NH, Desnoyers M. Clinical evaluation of a nutraceutical, carprofen and meloxicam for the treatment of dogs with osteoarthritis. *Vet Rec*. (2003) 152:323–29. doi: 10.1136/vr.152.11.323
48. Sadigh G, Parker R, Kelly AM, Cronin P. How to write a Critically Appraised Topic (CAT). *Acad Radiol*. (2012) 19:872–88.doi: 10.1016/j.acra.2012.02.005
49. Grindlay DJC, Dean RS, Christopher MM, Brennan ML. A survey of the awareness, knowledge, policies and view of veterinary journal editors-in-chief on reporting guidelines for publication of research. *BMC Vet Res*. (2014) 10:10. doi: 10.1186/1746-6148-10-10
50. Wareham KJ, Hyde RM, Grindlay D, Brennan ML, Dean RS. Sponsorship bias and quality of randomised controlled trials in veterinary medicine. *BMC Vet Res*. (2017) 13:234. doi: 10.1186/s12917-017-1146-9
51. Wareham KJ, Hyde RM, Grindlay D, Brennan ML, Dean RS. Sample size and number of outcome measures of veterinary randomised controlled trials of pharmaceutical interventions funded by different sources, a cross-sectional study. *BMC Vet Res*. (2017) 13:295. doi: 10.1186/s12917-017-1207-0
52. Arlt S, Heuwieser W. The staircase of evidence–a new metaphor displaying the core principles of EBVM. *Vet Evid*. (2016) 1:1. doi: 10.18849/ve.v1i1.18
53. Phillips B. *The Crumbling of the Pyramid of Evidence*. BMJ Blogs (2014). Available online at: http://blogs.bmj.com/adc/2014/11/03/the-crumbling-of-the-pyramid-of-evidence/ (accessed May 27, 2020).
54. Yetley EA, MacFarlane AJ, Greene-Finestone LS, Garza C, Ard JD, Atkinson SA, et al. Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints: report from a joint US-/Canadian-sponsored working group. *Am J Clin Nutr*. (2017) 105:249S–85. doi: 10.3945/ajcn.116.139097
55. O'Connor A. Is the simplicity of the evidence pyramid actually detrimental for understanding evidence? *Vet Evid*. (2017) 2:1. doi: 10.18849/ve.v2i1.100
56. Veterinary Evidence. *Guidance for Writing the Clinical Bottom Line*. London: RCVS Knowledge (2020). Available online at: https://veterinaryevidence.org/index.php/ve/writing-the-clinical-bottom-line (accessed May 27, 2020).
57. Cullen, JN, Sargeant, JM, Makielski, KM, O'Connor AM. The case-control design in veterinary sciences: a survey. *Prev Vet Med*. (2016) 134:179–87. doi: 10.1016/j.prevetmed.2016.10.008
58. Sargeant, JM, O'Connor, AM, Cullen, JN, Makielski KM, Jones-Bitton A. What's in a name? The incorrect use of case series as a study design label in studies involving dogs and cats. *J Vet Intern Med*. (2017) 31:1035–1042. doi: 10.1111/jvim.14741
59. EBVM Learning. *Appraise*. London: RCVS Knowledge (2015). Available online at: http://www.ebvmlearning.org/appraise/ (accessed May 27, 2020).
60. RCVS Knowledge. *EBVM Toolkit 4: What Type of Study Is It?* London: RCVS Knowledge (2015). Available online at: https://knowledge.rcvs.org.uk/

document-library/ebvm-toolkit-4-what-type-of-study-is-it/ (accessed May 27, 2020).
61. CEBM *Study Designs*. Oxford: Centre for Evidence-based Medicine (2020). Available online at: https://www.cebm.net/2014/04/study-designs/ (accessed May 27, 2020).
62. CEBM. *Critical Appraisal Tools*. Oxford: Centre for Evidence-Based Medicine (2020). Available online at: https://www.cebm.net/2014/06/critical-appraisal/ (accessed May 27, 2020).
63. CEVM. *Resources*. Nottingham: Centre for Evidence-based Veterinary Medicine (2020). Available online at: https://www.nottingham.ac.uk/cevm/resources/our-resources.aspx (accessed May 27, 2020).
64. Petrie A, Sabin C. *Medical Statistics At A Glance*. 3rd ed. Oxford: Wiley-Blackwell (2009).
65. Petrie A, Watson P. *Statistics for Veterinary and Animal Science*.3rd ed. Oxford: Wiley-Blackwell (2013).
66. Perera R, Heneghan C, Badenoch D. *Statistics Toolkit*. Oxford: Blackwell Publishing (2008).
67. CEVM. *Training Opportunities*. Nottingham: Centre for Evidence-based Veterinary Medicine (2020). Available online at: https://www.nottingham.ac.uk/cevm/training-opportunities/training-opportunities.aspx (accessed May 27, 2020).
68. Equine Veterinary Education. *Author Guidelines*. Fordham (2020). Available online at: https://beva.onlinelibrary.wiley.com/hub/journal/20423292/homepage/forauthors.html (accessed May 27, 2020).
69. Dawes M. Critically appraised topics and evidence-based medicine journals. *Singapore Med J*. (2005) 46:442–9.
70. Dong P, Wong LL, Ng S, Loh M, Mondry A. Quantitative evaluation of recall and precision of CAT crawler, a search engine specialized on retrieval of critically appraised topics. *BMC Med Inform Decis Mak*. (2004) 4:21. doi: 10.1186/1472-6947-4-21
71. Mackway-Jones K, Carley SD, Morton RJ, Donnan S. The best evidence topic report: a modified CAT for summarising the available evidence in emergency medicine. *Emerg Med J*. (1998) 15:222–6. doi: 10.1136/emj.15.4.222
72. Sparkes AH, Bessant C, Cope K, Ellis SLH, Finka L, Halls V, et al. ISFM Guidelines on population management and welfare of unowned domestic cats (Felis catus). *J Feline Med Surg*. (2013) 15:811–7. doi: 10.1177/1098612X13500431
73. Kasch C, Haimerl P, Heuwieser W, Arlt S. Evaluation of a CAT database and expert appraisal of CATs developed by students. *J Vet Med Educ*. (2017) 44:676–85. doi: 10.3138/jvme.0416-083r1
74. Greenhalgh T. Of lamp posts, keys, and fabled drunkards: a perspectival tale of 4 guidelines. *J Eval Clin Pract*. (2018). 24:1132–8. doi: 10.1111/jep.12925
75. CKS. *Clinical Knowledge Summaries*. London: National Institute for Health and Care Excellence (2020). Available online at: https://cks.nice.org.uk/#?char=A. (accessed May 27, 2020).
76. University of Glasgow. *Advanced Practice in Veterinary Nursing MSc*. Glasgow: University of Glasgow (2018). Available online at: https://online.gla.ac.uk/programme/university-of-glasgow-advanced-practice-in-veterinary-nursing-master-1544700394216 (accessed May 27, 2020).
77. University of Liverpool. *Certificate in Advanced Veterinary Practice (CertAVP) Module List*. Liverpool: University of Liverpool (2020). Available online at: https://www.liverpool.ac.uk/vets/cpd/certavp/module-list/ (accessed May 27, 2020).
78. RCVS Knowledge. *Veterinary Evidence Student Awards*. London: RCVS Knowledge (2020). Available online at: https://knowledge.rcvs.org.uk/grants/available-grants/veterinary-evidence-student-awards/ (accessed May 27, 2020).
79. Wootton CI, Cartwright RPE, Manning P, Williams HC. Should isotretinoin be stopped prior to surgery? a critically appraised topic. *Br J Dermatol*. (2014) 170:239–44. doi: 10.1111/bjd.12761
80. Belshaw Z. What is your client thinking and why should you care? *Vet Rec*. (2017) 181:517. doi: 10.1136/vr.j5234

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

REVIEW
published: 07 July 2020
doi: 10.3389/fvets.2020.00339

An Introductory Framework for Choosing Spatiotemporal Analytical Tools in Population-Level Eco-Epidemiological Research

Kaushi S. T. Kanankege[1], Julio Alvarez[2], Lin Zhang[3] and Andres M. Perez[1]*

[1] Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, Saint Paul, MN, United States, [2] Departamento de Sanidad Animal, Centro de Vigilancia Sanitaria Veterinaria (VISAVET), Facultad de Veterinaria, Universidad Complutense, Madrid, Spain, [3] Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, United States

Spatiotemporal visualization and analytical tools (SATs) are increasingly being applied to risk-based surveillance/monitoring of adverse health events affecting humans, animals, and ecosystems. Different disciplines use diverse SATs to address similar research questions. The juxtaposition of these diverse techniques provides a list of options for researchers who are new to population-level spatial eco-epidemiology. Here, we are conducting a narrative review to provide an overview of the multiple available SATs, and introducing a framework for choosing among them when addressing common research questions across disciplines. The framework is comprised of three stages: (a) pre-hypothesis testing stage, in which hypotheses regarding the spatial dependence of events are generated; (b) primary hypothesis testing stage, in which the existence of spatial dependence and patterns are tested; and (c) secondary-hypothesis testing and spatial modeling stage, in which predictions and inferences were made based on the identified spatial dependences and associated covariates. In this step-wise process, six key research questions are formulated, and the answers to those questions should lead researchers to select one or more methods from four broad categories of SATs: (T1) visualization and descriptive analysis; (T2) spatial/spatiotemporal dependence and pattern recognition; (T3) spatial smoothing and interpolation; and (T4) geographic correlation studies (i.e., spatial modeling and regression). The SATs described here include both those used for decades and also other relatively new tools. Through this framework review, we intend to facilitate the choice among available SATs and promote their interdisciplinary use to support improving human, animal, and ecosystem health.

Keywords: geographical/spatial analysis, geostatistics, epidemiology, disease mapping, framework

OPEN ACCESS

Edited by:
Flavie Vial,
Animal and Plant Health Agency,
United Kingdom

Reviewed by:
Kayoko Shioda,
Yale University, United States
Maysa Pellizzaro,
Universidade Federal da Bahia, Brazil
Alessandra Carioli,
University of Southampton,
United Kingdom

****Correspondence:***
Kaushi S. T. Kanankege
kanan009@umn.edu

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *04 February 2020*
Accepted: *15 May 2020*
Published: *07 July 2020*

Citation:
Kanankege KST, Alvarez J, Zhang L and Perez AM (2020) An Introductory Framework for Choosing Spatiotemporal Analytical Tools in Population-Level Eco-Epidemiological Research. Front. Vet. Sci. 7:339. doi: 10.3389/fvets.2020.00339

SPATIAL EPIDEMIOLOGY

Spatial epidemiology is defined as "the description and analysis of geographic variations in disease with respect to demographic, environmental, behavioral, socioeconomic, and infectious risk factors" (1). The importance of understanding the interplay between genetic, population, and environmental factors, and temporal characteristics of diseases in relation to space (2–4) has provided a set of powerful reasons to further develop the field of spatial epidemiology. The integration of epidemiological concepts, spatial analysis, geographic information system (GIS), and

statistics leads to the accomplishment of the objectives of spatial epidemiology in understanding and modeling spatiotemporally explicit health risks (5–10). Essentially, geostatistics was originated in fields of geoscience, and the use of geostatistics on health data is synonymously referred to as "medical/health geography" or "spatial/geographical epidemiology" (11, 12).

The poster child of spatiotemporal epidemiological studies is Dr. John Snow's map of cholera deaths in Soho, London, in 1854 (13, 14). Dr. Snow used the map to support his theory that disease was associated with contaminated water, contrary to the popular belief at the time that it was airborne (14). Dr. Snow's classic work is an early example of how spatial epidemiological methods may support improving the quality of epidemiological investigations, eventually providing risk estimates in a timely manner to support decision and policy in preventive and control measures (15–17). Traditionally, spatial epidemiology focused on two major concepts: (a) mapping and spatial pattern analysis, such as cluster analysis, to determine visual and geographical relational cues (pre-hypothetical stages of research), and (b) using ecologic approaches to recognize etiologic clues of disease spread and explanatory factors (hypothesis-driven research) (18). However, the emergence of a large variety of tools and methods over the last decades has made the landscape of spatiotemporal epidemiological tools quite complex, challenging researches ability to identify the analytical approaches most suitable for their needs.

SPATIOTEMPORAL VISUALIZATION AND ANALYTICAL TOOLS (SATs)

A plethora of SATs, especially geostatistical tools, have been published and used in the field of spatial epidemiology (15, 19). However, for a beginner in spatial eco-epidemiology, selecting an appropriate analytical tool is often a challenging decision. Different disciplines, including epidemiology, econometrics, and ecology, use different SATs to address similar research questions (20–23). Juxtaposing these diverse techniques may support an interdisciplinary approach of shared knowledge while providing a list of options for researchers. The choice of SATs depends on a variety of factors/criteria. The majority of the published reviews and books on SATs are focused on describing the features of the tools/methods and do not guide a beginner researcher through the options to consider when choosing a spatial eco-epidemiological analysis. The objective of the paper here was to suggest a framework that facilitates choosing SATs which enables the researchers to analyze existing epidemiological data, draw inferences, and plan future research in spatiotemporal epidemiology.

DATA USED IN SPATIOTEMPORAL ANALYSIS

The types of spatial data that can be used in epidemiology to represent the distribution of diseases and adverse events in space include (1) point-referenced data (presence and absence of the disease or number of animals at each farm location), (2) point-pattern data (presence of the disease: where the disease occurrence itself is random giving rise to a "spatial point process"), and (3) areal data or "lattice data" (number of disease cases aggregated by an administrative division such as counties) (19, 24). The first case is often referred to as "geocoded" or "geostatistical" data (19). The point-referenced data and areal data may be of binary, count, or continuous in nature. The key difference between point-referenced and point-pattern data is that the former has a set of pre-known locations from which a certain value for a given variable was observed, whereas in the latter the events are assumed to have a stochastic or random nature (19). Therefore, in point-pattern data both the location and the observation of the disease themselves are random or stochastic. While the term "lattice data" may lead to the assumption that the areal units are regular shaped grids, in practice most areal data are summarized over irregular lattice such as administrative divisions. Reduced spatial explicitness may lead to aggregation of the events by administrative divisions and non-availability of the temporal details would limit the researcher to use purely spatial tools for the analysis.

While disease status data are the primary focus, epidemiological studies often look into association of the disease with underlying risk factors, such as human population density, air pollution parameters, temperature, precipitation, or soil pH among many other possible examples, which vary continuously over the space. These variables that are usable on GIS platforms are available from various data base sources in the form of point-referenced observations, polygon maps, or gridded i.e., "raster" maps. WorldClim [www.worldclim.org; (25, 26)] and LandScan Global Population Database (27) are examples of such data sources. The relevant value of these continuous variables, at each location where the disease status has been determined, can be extracted and used for further analysis, i.e., point-referenced data (19). The availability of exact location details and the time of the case supports more spatiotemporally explicit and reliable analysis. Unless specified as applicable to a particular type of data only, SATs described here are suitable to be used point-pattern, point-referenced, or areal data. It is important to notice that under certain circumstances the data types can be converted from one form to another. Point-referenced data can be summarized and represented by administrative divisions (i.e., polygon data). For example, point-referenced data representing 10 different farm locations recorded with a disease can be represented as 10 cases with in the county. Similarly, disaggregation of areal data with certain assumptions, such as density dependent disaggregation (28), is possible. Representing the area by the centroid of each polygon, thus, converting areal data into a point-referenced format, which, of course, is a simplification of the analysis that may be acceptable only under certain circumstances.

A FRAMEWORK FOR CHOOSING SPATIOTEMPORAL EPIDEMIOLOGICAL TOOLS

Here, we are suggesting a framework for choosing SATs (**Figure 1**). The framework is classified into three stages: (a) pre-hypothesis testing/hypothesis generating stage; (b) primary

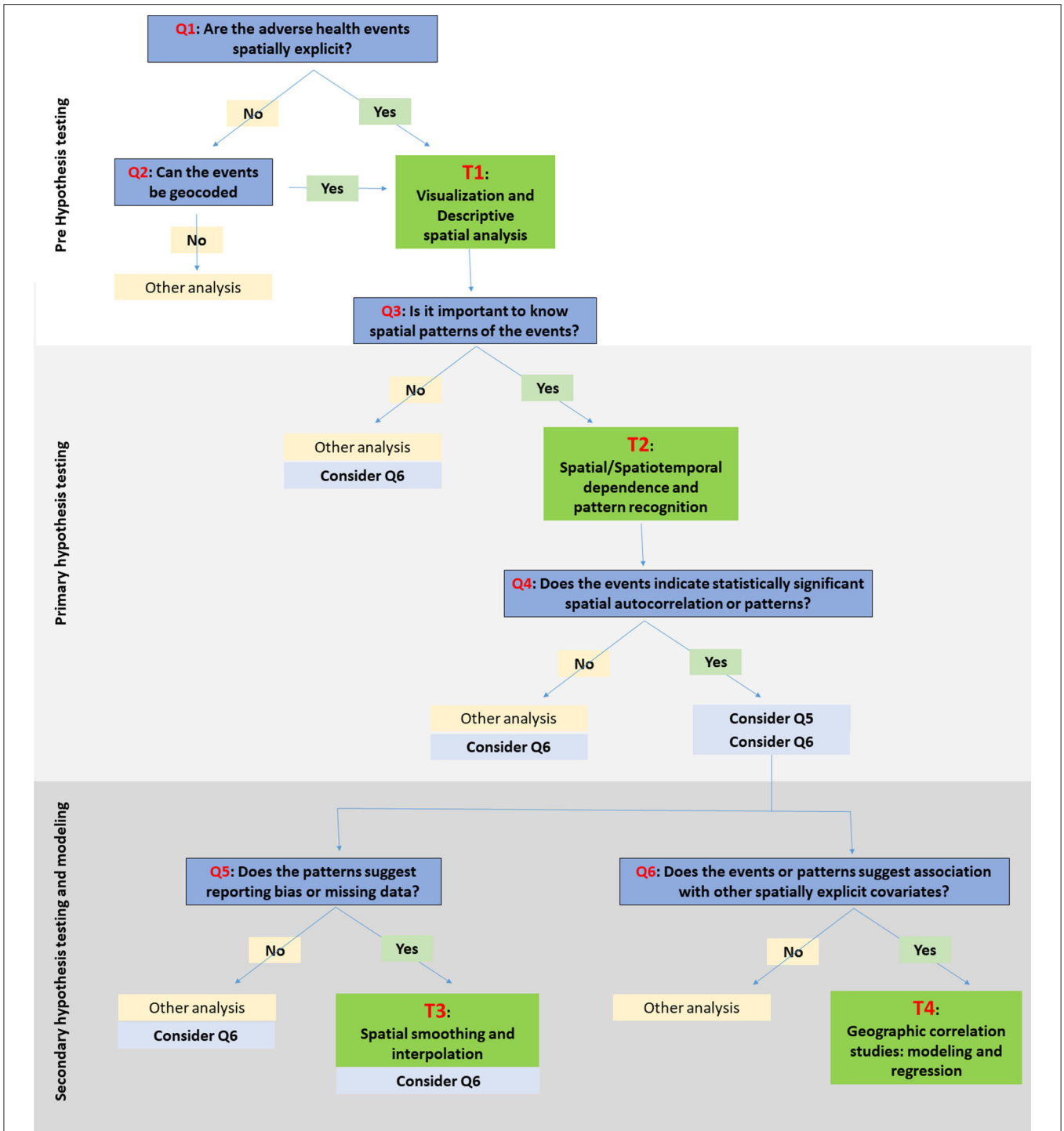


FIGURE 1 | Schematic illustration of a framework for choosing spatiotemporal visualization and analytical tools (SATs). The research questions/objectives are identified with Q1:Q6. The specific SATs under the relevant categories, i.e., T1:T4, are listed in **Table 1**.

hypothesis testing stage; and (c) secondary-hypothesis testing and spatial modeling stage where the predictions and inferences are made. The primary hypothesis refers to the existence of spatial dependence and spatial patterns in the distribution of adverse health events, while the secondary hypotheses involve the association of the events with risk factors/covariates. The different types of SAT are broadly classified into four categories: (T1) visualization and descriptive analysis; (T2)

spatial/Spatiotemporal dependence and pattern recognition; (T3) spatial smoothing and interpolation; and (T4) spatial correlation studies: modeling and regression. The types of data primarily applicable with different SATs are listed under T1:T4. The framework seeks to suggest a suitable category of the SAT among the four, based on the stage of the research question. The types of SAT that are commonly used in epidemiological studies are listed under each category (T1:T4) in **Table 1** and discussed briefly below. The usage of tools are further discussed in relation to one example case study. It is important to note, however, that this is not a systematic review on the existing SATs, and that the classification used here is, somewhat, arbitrary, given the subjective nature of the problem. This contribution of a narrative review, while not an exhaustive description of SATs, intends to provide a short guide to introductory-level population and ecological scientists on commonly used tools and encourage the users to explore the diverse algorithms for more informed conclusions. Detailed reviews on SATs can be found elsewhere (6, 7, 10, 23, 138), as well as, a glossary of commonly used terms and their definitions in spatial epidemiology is found in Rezaeian et al. (11).

COMMONLY USED SPATIOTEMPORAL VISUALIZATION AND ANALYTICAL TOOLS (SATs)

T1 Tools for Visualization and Descriptive Analysis

Spatial data visualization is one of the key steps in understanding and generating hypotheses on the spatial distribution of events. Global Navigation Satellite Systems (GNSS), such as Global Positioning System (GPS); Global Navigation Satellite System (GLONASS); Galileo; Navigation Indian Constellation (NavIC); and BeiDou provide the ability to position the exact geospatial locations during the data collection phase. In the absence of GNSS based data, geocoding plays a major role to generate spatially explicit databases (29, 30). In addition to the visualization, description of the extent of spatial distribution by means of size, shape, and directionality of the spread supports understanding the extent of the adverse health/environmental effect. Descriptive analysis using T1 tools may support planning primary interventions including assigning vaccine or surveillance buffer zones and recognizing the distance to closest epidemiologically important features.

GIS is a system which enables capturing, storing, visualizing, and analyzing spatially explicit or "georeferenced" data to cartographic projections (31, 139). The true value of the ability to place data or measurements on a map, either as discrete events using its exact location (i.e., point-referenced data) or as continuous data by regular grids (i.e., raster data), is the ability to assess possible relationships within the data. GIS technology makes it technically feasible to integrate large amounts of data collected from different sources into a single georeferenced map/model for analysis. Therefore, GIS plays a major role in the spatial analysis as a platform which facilitates bringing data and analytical techniques together. The key analytical tools are listed under T2:T4.

T2 Tools for Spatial/Spatiotemporal Dependence and Pattern Recognition

Measures of Spatial Autocorrelation

According to Walter Tobler's First Law of Geography, "everything is related to everything else, but near things are more related than distant things (140)." This phenomenon, otherwise known as spatial autocorrelation or spatial dependence, is a key component of spatial epidemiology. The majority of the T2 techniques are focused on determining the extent to which data are spatially autocorrelated and performing hypothesis tests after accounting for spatial autocorrelation (141). Assumptions involved in the analytics include the spatial stationarity, isotropic spatial autocorrelation, and spatial continuity (141). In simpler terms these assumptions imply that events (infectious diseases in animals for example) of the considered spatial process are homogeneously distributed across the region regardless of geographical directions or barriers. However, understanding the violations of these assumptions, i.e., detecting patterns of non-stationarity or anisotropy, is paired with the descriptive analytics (32). Moran's I (37), Geary's C (38), Mantel test (39), and Getis Ord (40, 41), which often referred to as "global spatial autocorrelation indices" (142) are the commonly used techniques to measure spatial autocorrelation.

Measurement of spatial heterogeneity, i.e., uneven distribution of the populations and risk factors across the geographical space, is another important component for understanding the disease process. Spatial heterogeneity measures could be either (1) local where we measure whether an attribute at one site is different from its surrounding or (2) stratified where the attributes are stratified within strata, such as Agro-ecological zones or land use categories in which the spatial variance between strata was measured. An example of local measures of spatial heterogeneity is Getis Ord Gi* [i.e., hot-pot/cold spot analysis; (40, 41)]. Other techniques such as G-statistics are increasingly available facilitating the measurement of stratified spatial heterogeneity (51). The indices of spatial heterogeneity provide opportunity to quantitatively measure the differences and compare the landscape patterns of populations and risk factors.

Spatial Cluster Analysis

A spatial cluster is an excess of events or measurements in certain areas in geographic space, compared to the null expectation of complete spatial randomness (143). The cluster analysis is generally aimed at detecting if there is any clustering in the spatial data (i.e., Global cluster analysis), and detecting and locating the clusters (local cluster analysis and focused cluster analysis). In general, the cluster analysis provides information about the cluster morphology, including the magnitude of the excess/deficit feature, geographic size, shape, and the locations of spatial clusters.

Detecting first-order adjacencies such as Local Indicators of Spatial Autocorrelation (LISA) statistics (41, 50) and nearest-neighbors relationships such as used in Cuzick and Edward's

TABLE 1 | A summary of types of common spatial analytical tools and their purpose.

	Purpose	Measure	Commonly used techniques	D*	References
T1: Visualization and descriptive analysis	Transformation of locational information into geographic coordinates	Geocoding/ georeferencing	GIS based geocoding of street address, postal code, or administrative divisions	pp, pr, ar	(29–31)
T2: Spatial/ Spatiotemporal dependence and pattern recognition	Visualization and description of the size and shape of the spatial distribution	Exploratory spatial data analysis	Mean center	pp, pr, ar	(32)
			Median center		(32)
			Convex hull		(33)
			Standard deviation (weighted by attributes)		(32)
			Directional mean and variance		(34)
			Moran scatter plot		(35)
		Characterize nearby features	Features with in a distance band/buffer zone	pr, ar	(31, 36)
			Distance to feature		(31)
			Overlaying features		(31)
	Test whether there is spatial dependence in the event data	Spatial autocorrelation	Global Moran's I	pr, ar	(37)
			Geary's C		(38)
			Mantel test		(39)
			Geti's ord		(40, 41)
		Spatial autocorrelation among regression residuals	Moran's I test	pr, ar	(42, 43)
			Kelejian–Robinson test		(44, 45)
		Distance analysis	Nearest neighbor analysis		(46)
			Ripley's K		(47, 48)
			Distance matrices		(31)
	Measure the uneven distribution of the populations and risk factors	Local or stratified spatial heterogeneity	Getis Ord Gi*	pr, ar	(40, 41)
			K-means clustering		(49)
			Anselin's local Moran's I (L-Moran)		(50)
			Spatial stratified heterogeneity test		(51)
	Measure the spatial dependence while accounting for background population		Oden's Ipop	ar	[(52, 53); https://www.biomedware.com]
	Test whether there is any spatial trends	Testing for first-order effects	Trend analysis	pr, ar	(18, 54, 55)
	Test whether there is any spatial clustering in the data	Global cluster detection	Nearest neighbor test	pp, pr, ar	(46)
			Cuzick and Edward's test (case-control data)		(56)
			Local indicators of spatial association (LISA)		(50)
	Locate the clusters and the statistical significance of the clustering	Purely spatial local cluster detection	Spatial scan statistics		(57–59)
			Flexscan	ar	(60)
			Turnbull's test	pr, ar	(61)
			Besag and Newell's test		(62)
	Test whether there is space and time clustering in the data	Spatiotemporal cluster detection	Knox test	pp, pr, ar	(63)
			Mantel test		(39)
			Barton's test		(64)
			kth nearest neighbor test for time-space interaction		(65)
			Space-time permutation scan statistic		(66, 67)
			Edrer-Myers-Mantel test		(68, 69)
	Detect the direction of progression of an event over time	Spatiotemporal directionality	Spatiotemporal directionality test	pr, ar	[(53, 70); https://www.biomedware.com]
			Spatiotemporal anisotropy parameter		(71, 72)

(Continued)

TABLE 1 | Continued

	Purpose	Measure	Commonly used techniques	D*	References
T3: Spatial smoothing and interpolation	Quantifying spatial variations in event intensity: spatial point pattern (SPP) intensity	Density based point pattern recognition	Univariate Kernel density estimation (KDE)	pr	(73–75)
			Multidimensional KDE		(76, 77)
			Empirical Bayes smoothing (EBS)	ar	(78, 79)
	Smoothing and interpolation	Deterministic spatial interpolation	Thiessen (Voronoi) polygons	pr	(80)
			Neighborhood matrices		(31)
			Inverse Distance Estimation (IDW)		(32, 81, 82)
			Triangulated Irregular Network (TIN)		(83, 84)
			Headbang smoothing		(85–87)
		Spatial modeling with stochastic partial differential equations (SPDE)		pr	(88, 89)
		Geostatistical interpolation and spatial regression	Kriging	pr	(32, 90, 91)
			Spline regression models		(92)
			Trend Surface Interpolation		(93–96)
		Multivariate spatial interpolation	Co-kriging	pr	(32, 91, 97)
			Regression kriging		(98–100)
		Spatiotemporal interpolation	Space-time kriging	pr	(101, 102)
			Autoregressive spatial smoothing and temporal Spline smoothing		(103)
T4: Geographic correlation studies: modeling and regression	Estimate the probability of disease spread using explanatory variables	Regression at spatial units	Ordinary least square regression and test for spatial autocorrelation of residuals	pp, pr, ar	(42, 43, 45)
			Spatial lag model with independent variable representing neighbors		(104, 105)
		Spatial and spatiotemporal error autoregression models for areal data (When regression residuals have spatial autocorrelation)	Simultaneous autoregressive (SAR) models	pr, ar	(19, 24, 106)
			Geographically weighted regression (GWR)		(107, 108)
			Purely spatial: Conditional autoregressive (CAR) models		(19, 109, 110)
			Spatiotemporal CAR models		(111, 112)
			Two-stage space-time mixture modeling		(113)
			Latent structure models		(113–115)
		Spatial and spatiotemporal models for point-level data	Point process models with weighted sum approximation	pp	(116, 117)
			Conditional logistic model	pp, pr	(118, 119)
			Separable models for spatiotemporal data		(19)
			Non-separable models for spatiotemporal		(19)
	Measure the gravitation of adverse effects and the risk factors based on distance	Estimate most probable spatial interactions between entities	Gravity models	pr, ar	(120–123)
	Analysis of spatially explicit time-to-event data	Spatial survival models	Spatial cure rate model	pr	(124)
			Frailty models		(124)
	Estimate the probability of disease when the disease occurrence is correlated with environmental variables	Environmental/Ecological niche modeling	Maximum Entropy Ecological Niche modeling (Maxent)	pr	(125–127)
			Genetic Algorithm for Rule Set Production (GARP)		(128–130)
		Machine/statistical learning techniques	Random forest	pr	(131, 132)
			Generalized additive models (GAMs)		(133–135)
			Artificial neural networks (ANN)		(136, 137)

D Column represents the type of data primarily applicable on the set of tools, where, pp, point-pattern; pr, point-referenced; ar, areal data.*

(56) test can be considered as global cluster detection techniques. Most local cluster-detection techniques employ circular scanning windows, such as the scan statistic (58), Turnbull's test (61), and Besag and Newell's (62) test. In scan statistics, a circular scanning window of varying sizes that moves across the study area is used to compare the observed-to-expected ratio of the cases compared to the expected spatial randomness was calculated, and the windows that maximize this likelihood ratio were recognized as the most likely clusters (58). Some of these local cluster analyses such as scan statistics have been incorporated into widely used software such as SaTScan that enable temporal, spatial, and spatiotemporal cluster analysis in a user-friendly manner. However, it is essential to realize that spatial variation and hence cluster morphology is complex, and may not be well-described by the circular cluster window approaches (143, 144). Therefore, alternative approaches that are flexible for the cluster shape such as Flex scan (60), Upper Level Set scan statistics (145), and B-statistics (146) have been introduced. A detailed description on the spatial pattern recognition and cluster analytical techniques are found elsewhere (143). The performance of SATs designed to detect clusters can be highly sensitive to the level of aggregation of the data (147). Therefore, while the clusters detected based on point-pattern or point-referenced data are intuitive to interpret, the clusters of data aggregated at large areal units requires caution. Distance based assignment of the neighbors instead of considering shared borders between areal units has been suggested (147). Morris and Munasinghe (148) have offered a solution through a user defined computer algorithm that combines existing areal units, such as administrative divisions, into regions with populations large enough to diminish spurious variability in disease rates while limiting the loss in resolution.

T3 Tools for Spatial Smoothing and Interpolation

Spatial Smoothing Techniques

Many research studies on adverse health/environmental events apply spatial smoothing and interpolation techniques to improve estimation and for exploratory mapping of risk (149). There is a variety of smoothing techniques and they can be broadly categorized as global (the same function is applied to all the data points and predictions are made using the entire dataset) and local (the same function is applied to sub-sets of data points based on the neighborhood) smoothing techniques. Kernel smoothing, one of the widely used techniques, facilitates visualization of the intensity of events (73) while accounting for background spatial distribution of the population at risk (150), and generate tolerance contours (i.e., confidence regions) for which the relative risk of a disease is significantly high (74, 75). Kernel smoothing can be used to describe and visualize the intensity or the spatial relative risk of health threats. Smoothing techniques are used to reduce noise by shrinking values toward the adjacent observations and estimate the spatial trend, which is applicable to both homogenous and heterogeneous point processes (75, 151). In a heterogeneous point process in which the intensity of the spatially varying event varies within the study area, smoothing is used to increase accuracy of the estimation of the event intensity using either parametric or non-parametric methods (73–75). Spatial smoothing techniques use a moving weighted function to reduce the noise component, where the differences in the values on a surface are accentuated resulting in a spatially continuous map. Commonly used spatial smoothing techniques include kernel density estimation (KDE) [(73, 74, 152, 153)] and headbanging (85–87), which are considered as alternatives of detecting circumscribing clusters of varying shapes in lieu of circular clusters (74, 143). Empirical Bayes smoothing (EBS) is a specific case of spatial smoothing where the denominator i.e., varying population at risk over the map is used as a measure of the confidence in risk estimates. Therefore, the confidence of estimates are higher in highly populated areas, whereas, the estimates of relative risk would have high margins of error in the less populated areas (79). For example, if two counties have same the standardized incidence ratio (SIR) but have different population sizes, the confidence of EBS estimates would be higher for the county with a larger population size.

Spatial Interpolation Techniques

Spatial interpolation techniques are used to estimate or predict values at unknown locations using available/known data points (32). These tools can be broadly categorized as deterministic (they use the extent of similarity or distance to create the surface using measured points) and geostatistical (they use the statistical properties of the measured points to create the interpolated surface) interpolations. The resulting interpolated surfaces i.e., statistical surfaces are raster layers and often can be considered as risk maps in epidemiological analyses. There are multiple spatial interpolation techniques including Inverse distance estimation (IDW) (81), Triangulated Irregular Network (TIN) (5, 83), Kriging as well as its variations such as Co-kriging (32), and Trend Surface Interpolation (93–96) are among the commonly used techniques. TIN represents the surface by a set of contiguous and non-overlapping triangles connecting the original data points and allows construction of 3-dimensional surfaces based on a secondary variable of a researcher's choice, which, for example, the prevalence of a disease in a farm location. A review by Li and Heap (84) summarizes and compares several interpolation methods used in environmental sciences that are highly applicable in eco-epidemiological studies as well.

Geostatistical interpolation, such as kriging can be understood as a two-step process, where, step 1 is fitting the spatial variogram or likelihood for the data observed at the sampled points; and step 2 involves the interpolation of values for unsampled points or blocks using the weights derived from this covariance structure (32). In situations in which disease events are biased or undersampled, co-kriging can be used to enhance the accuracy of the estimation using a highly sampled auxiliary variable (154). For example, when invasive species detected at lakes are underreported, but the known invasions are highly correlated with the visitors/boater traffic in-and-out of the lakes and data are available for this variable, boater traffic network may use as an auxiliary variable to determine the lakes that are likely to be invaded (155). Trend surface interpolation facilitates mapping variables while allowing for the local fluctuations. Therefore, trend surface analysis may reflect the regional distribution,

trend, and the local variation of the mapped disease (156, 157). Interpolation techniques, their model assumptions, and usage are discussed extensively, elsewhere (32, 96).

Spatiotemporal interpolation techniques are used to predict variables in-between and beyond observation times (101, 102). In space-time kriging, the spatial, temporal, and spatiotemporal dependence structures are modeled using spatiotemporal variograms (102). Modeling the spatial and temporal components independently is one of the drawbacks in most of the spatiotemporal interpolation techniques (158). A detailed discussion on the spatiotemporal interpolation techniques used in the environmental modeling is found elsewhere (158). Recent developments including spatial modeling with stochastic partial differential equations (SPDE) have further improved spatial and spatiotemporal smoothing using Bayesian inference (88, 89).

T4 Tools for Geographical Correlation Studies: Modeling and Regression

Spatial Regression Models

In geographic correlation studies in epidemiology, spatial regression analysis is commonly used to examine the effects of certain risk factors/covariates on disease incidence while accounting for the spatial autocorrelation/dependence (19, 104, 159–161). Spatial dependence is incorporated into the model specifications typically using a spatial lag term or spatial error autorregression models [i.e., assigning autoregression terms for regression residuals; (104, 160)]. This is because the standard regression models assume that observations are independent, an assumption that is not met when spatially dependent data are analyzed. Fitting regression models while assigning a variable to represent the neighbor effect is one way of modeling the spatial dependence. For example in spatial lag model in which we assume that disease status in at one location is affected by the disease status at the nearby locations, a "lag" term, which is a specification of disease status at nearby locations, is included in the regression, and its coefficient and *p*-value are interpreted as for the independent variables (104). Both Frequentist and Bayesian spatial regression techniques have been extensively used in epidemiological analyses. Spatial regression models vary by their computational complexity, capacity of capturing spatial heterogeneity, and the quantification of uncertainty associated with parameter estimates (161).

Spatial error autoregressive models for discrete/areal data include: Simultaneous autoregressive (SAR) models (19, 24, 106, 162), Geographically weighted regression (163), and Conditional autoregressive models (CAR) with neighborhood structures defined based on Besag, York, and Mollie (BYM) model or Leurox (109, 110). Defining the neighbors for areal data is done based on contiguity including first-order contiguity (i.e., presence of shared borders between polygons such as adjacent counties); graph-based contiguity (i.e., based on defined algorithms such as nearest-neighbor graphs); or distance-based contiguity [i.e., neighbors within 10 km; (45)]. Due to sampling and reporting variabilities of disease incidences and risk factors, borrowing strength from neighboring regions to get more reliable estimates is the motivation behind these spatially dependent regression models (e.g., closer neighbors might receive higher weights). This strategy of borrowing information from neighbors is applicable in autoregressive models, where the spatial or spatiotemporal structure is modeled via sets of autocorrelated random effects (19, 109, 164).

In addition to accounting for the spatial dependency, multiple spatiotemporal regression models have been used in epidemiological studies that enable the researchers to analyze the influence of spatial and temporal dependence of disease events and risk factors (19, 165). Detailed descriptions on spatial and spatiotemporal autoregressive models can be found elsewhere (19, 165). For example, latent structure models which accounts for the heterogeneity or the discontinuity in risk surface such that homogenous areas can be grouped together while discriminating for the risk levels (114).

When the events are recorded as point-referenced data from locations within a continuous spatial domain, such as by households or animal farms in a certain area, the binary outcome that the adverse event occurs in each location is assumed to have an underlying continuous spatial process. Spatial processes with binary outcomes are usually modeled by spatial logistic or probit regression models. Assigning the spatial dependence and neighbors in spatial process is complicated. This is because point-referenced spatial data often come as multivariate measurements at each location and we anticipate dependence between measurements both at a particular location as well as across locations. For example presence of a certain animal disease in a farm is correlated with the farms own characteristics including number of animals and management practices, as well as the presence of neighboring farms. Separable and non-separable spatiotemporal regression models are commonly used to model spatial point processes (19, 166).

Environmental/Ecological Models

Ecological niche modeling (ENM) approaches are widely used to characterize the complexity and heterogeneity of the landscapes in research related to epidemiologically relevant vector and parasite-reservoir distributions (167, 168). In addition to the characterization of the areas where disease is distributed, ENM is used to identify potential distributional areas in response to the likely geographic shifts in distributional areas of species or phenomena under scenarios of climate change or changing land use (169). Genetic Algorithm for Rule Set Production (GARP) (129, 130); Maximum Entropy Ecological Niche modeling (Maxent) (125, 126); and Machine/statistical Learning Techniques such as random forest (131, 132) and artificial neural networks (ANN) (136, 137) are the commonly used algorithms in epidemiology. Most ENM studies use presence-only data for the analyses. Further details regarding GARP, Maxent, and other ENM algorithms are found elsewhere [(125, 126, 128, 129)]. Additionally, hybrid methods that are bringing together multiple tools are being used in several disciplines to improve estimation and prediction abilities in spatial analysis.

EVALUATING THE PERFORMANCE OF SPATIOTEMPORAL ANALYTICAL TOOLS

Model Performance Indicators

Evaluating model performance is important when choosing between similar SATs (Especially those listed under T3 and T4). These measures include correct classification rate (CCR) (170), model sensitivity and specificity (i.e., the number of correctly classified cases) and area under the receiver operating characteristics (ROC) curve (170, 171). The sensitivity of a spatial model in disease mapping can be defined as the model's ability to correctly predict high-risk areas/locations, whereas, the specificity of the model would be its ability to correctly identify low-risk areas/locations. Error and accuracy measures, such as root mean squared error (RMSE), are also used to measure how wrong the resultant model estimates can be (138). Similarly, penalized-likelihood criteria for comparing models including Akaike information criterion (AIC) (172), Bayesian information criterion (BIC) (173), Deviance information criterio (DIC) (174, 175), and Watanabe-Akaike information criterion (WAIC) (176) are used in regression models as relative measures to compare between models and evaluate goodness of fit with penalty on model complexity. Further reading on the choice of model selection criterion is found elsewhere (177, 178).

Model Validation Techniques

The SATs, especially the predictive modeling-and correlation models (listed under T3 and T4 of **Table 1**), are evaluated for their performance because the predictions would have no merit if the accuracy of the models cannot be assessed using independent data (138, 170, 179). A variety of techniques are available to validate the SATs (Listed under T3 and T4 of **Table 1**). Data partitioning techniques such as bootstrapping (180, 181), randomization (182), prospective sampling (182, 183), and k-fold partitioning (184, 185), leave-one-out cross-validation (138) are commonly used to determine training and testing datasets for model validations.

Cross validation, i.e., partitioning the data into several subsets and each fitting the model excluding one subset and validating the fitted model's ability to correctly predict the risk areas using the excluded subset of data, is one of the common practices in spatial model validation (138, 185). This includes dividing the data over space or time. For example, if the incident data are from 2000 through 2018, fitting model using early data/incidents and validation of the model predictions using recent events is considered an approach of temporal cross validation. Temporal cross validation is also achieved through the prospective sampling where new cases are evaluated against already built models from a different region or from a different time (170). A review by Anselin (179) discuss model validation techniques used in spatial econometrics in relation to the statistical validity of the models. The model fitting concerns related to theory, hypothesis testing, choice of criteria, and practical considerations are discussed under this criteria of model validations (179).

AVAILABLE SOFTWARE TOOLS FACILITATING SAT

Multiple free and proprietary software tools are available facilitating the spatiotemporal analytical studies. However, there is no quality control over to assess the accuracy, reliability, and sustainability of the majority of those non-proprietary software. Some software, such as SaTScan™ (https://www.satscan.org) and ArcGIS (https://geocode.arcgis.com), have become successful commercial products that are widely in use (7, 186), while others are underutilized due to less popularity and irregular maintenance. Sustainability and maintenance of these software is essential when incorporating these software based eco-epidemiological analyses into surveillance or intervention measures. An overview of the spatial data analytical software is found elsewhere (186).

Geocoding can be implemented using either commercial GIS software or online that are developed by governmental (Ex. USGS map locator: https://store.usgs.gov/map-locator), private (ArcGIS Online Geocoding Service by Esri (https://geocode.arcgis.com/arcgis/); QGIS Geocoding Plugins (https://plugins.qgis.org/plugins/GeoCoding/); Geocoding using Google maps (https://cloud.google.com/maps-platform), or through educational organizations (e.g., TAMU Geo coding Services of the University of Texas A&M: http://geoservices.tamu.edu/). Similarly, Python based geocoding using open or commercial spatial data repositories and spatial database management systems such as Google geocoding application programming interface (API) and improving the capacity of spatial computing is a field in developing (187). These software and tools enable both batch geocoding where multiple addresses are submitted at once for geocoding, and reverse geocoding, i.e., determining the nearest street address based on given coordinates.

The commonly used user-friendly software in the spatiotemporal analysis that are capable of performing the descriptive analysis, spatial pattern recognition, smoothing/interpolation, and/or spatial modeling are ArcGIS (188), QGIS (189), GRASS (190), GeoDa [(191); http://geodacenter.github.io/index.html], Clusterseer [(53); https://www.biomedware.com/], SaTScan (http://www.satscan.org/version 9.6), and CrimeStat (192). Similarly, there are multiple toolboxes relevant to spatiotemporal analysis that can be used through following software: R statistical software (193), SAS (194) (SAS/STAT® software), STATA (195), and Matlab (Matlab: https://www.mathworks.com)[1]. platforms that are specifically developed for handling geospatial analysis. Some of the advanced statistical software packages enables performing both frequentist and Bayesian spatial analyses. For example, the R package "spatialreg" (196, 197) enables performing frequentist spatial error models including CAR models (listed under T4), while R packages "CARBayes" (198), "CARBayesST" (165), and "R-INLA" [(88); www.r-inla.org; (199)] enables fitting Bayesian CAR models using Markov Chain Monte Carlo (MCMC)

[1] *MATLAB* and Statistics Toolbox TM Release 2018a. Natick, MA: The MathWorks, Inc.

or Integrated Nested Laplace approximation (INLA) based estimation of the posterior distributions, respectively.

HOW TO USE THE FRAMEWORK TO CHOOSE SAT: AN EXAMPLE

While we have introduced a framework and a categorization of commonly used SATS, it is important to note that the choice of the SATs is entirely a researcher-driven decision. There are certain factors/criteria associated with the decision of choosing one method over the other. The factors include: (1) characteristics of the disease/adverse event; (2) study design; (3) spatial explicitness of data; (4) data quality and availability; (5) research question and hypothesis; (6) stakeholder involvement; and (7) existence of resources, policy, and regulations for the mitigation of events (200). These factors influences the six questions (Q1:Q6) illustrated in the framework (**Figure 1**).

For example, assume a researcher is interested in understanding epidemiological characteristics of natural Anthrax in animal populations and intends to use that information to plan a surveillance/vaccination program in an endemic area. Let us assume that the final output the researcher intends to have is a criteria to define zoning distances for ring vaccination or surveillance when at least one Anthrax case is reported. Firstly, understanding the extent of spread and duration of previous Anthrax outbreaks would play a major role when determining this surveillance/vaccination radii. Secondly, understanding the association between the epidemiological drivers of the disease and the characteristics of susceptible population would be of importance when planning an area-based surveillance/vaccination program.

At the pre-hypothesis stage of the framework (**Supplementary Figure 1**), answering questions Q1 and Q2 would guide the researcher to use T1 tools and obtain a spatially explicit data set that is ready for further spatial analysis. Anthrax, caused by a spore-forming bacterium *Bacillus anthracis*, is characterized by the prolonged survival of the spores on soil and wide range of hosts including wildlife, livestock, and human (201, 202). Therefore, the observational study designs on Anthrax are likely to be retrospective based on reported cases (203). Given Anthrax is reportable to the animal and public health authorities, most likely type of data available would be point-referenced in nature (i.e., presence of the disease at farm locations or grazing lands). Although in rare situations, data may be available aggregated at administrative divisions due to privacy policy. If the coordinates of case locations are not recorded along with the case report, geocoding the locations based on the descriptions or farm addresses would be the initiating step.

Once geocoded, answering the Q3 and the use of SATs listed under T2 would facilitate the recognition of spatiotemporal dependence between the reported cases (i.e., the primary hypothesis testing stage). Given the prolonged survival of Anthrax spores in contaminated soils/environment, in addition to the initial testing for spatial dependence, understanding the spatiotemporal dependence and spatiotemporal directionality is the key to understand the extent of past spread of the disease. Testing whether there are space and time clustering in the data would facilitate determining any particular area/s with high relative risk for disease clusters at a specific time [i.e., disease hot-spots; (203)].

Once geocoded, the primary hypothesis testing stage of the framework and the T2 tools would facilitate the recognition of spatiotemporal dependence between the reported cases and determining any particular area with high relative risk for disease clusters [i.e., disease hot-spots; (203)]. Given the prolonged survival of the Anthrax spores conducting purely spatial and spatiotemporal dependence and directionality is the key to understand the extent of past spread of the disease. This spatiotemporal pattern detection may lead to the refinement of further research questions (Q4: Q6 of the framework) and secondary hypothesis testing using the SATs listed under T3 and T4 (**Supplementary Figure 1**).

Because the pathogen is invariably dependent upon the distribution of susceptible species and environmental characteristics such as soil pH, rain fall, and flood plains; the choice of predictive modeling using correlated environmental factors such as regression or ecological niche modeling (ENM) (204) is a suitable option to consider (i.e., tools under T4). However, it is important to recognize that the ideal analysis for a chronic disease like Anthrax would be spatiotemporal correlation models that enable incorporating temporal changes of both the disease and underlying environmental characteristics, in addition to space.

Once the range of cluster radii (T2 tools) and key epidemiologically important environmental factors by area (T4 tools) were identified, these two key pieces of information would facilitate informing the decisions of planning the ring vaccination/surveillance programs. For example, recognition of which areas are at high risk for Anthrax based on the models outputs from T4 tools, such as ENM (204), and the extent/cluster radii of past outbreaks using T2 tools would allow us to inform defining the minimum and maximum zoning distances for ring vaccination/surveillance.

ADVANTAGE, CHALLENGES, AND DRAWBACKS OF SATs

The framework provides an introductory guide for choosing SATs for eco-epidemiological studies. Use of SATs improves an eco-epidemiological investigation by adding precision, facilitating the comparison of distributions by means of quantitative criteria, and capturing risk factors and characteristics that are unlikely to be detected by visual inspection or analyzing data without the spatial component (6). Therefore, SAT outcomes, commonly represented as "risk maps," may serve as estimates of the effects of "real" exposures to human, animal, and environmental health threats and facilitate recognizing the effect size at more vulnerable locations and time periods.

Common weaknesses associated with the spatial analysis and risk mapping are related to shortcomings in the accuracy of data,

choices of mapping and projections, choice of the analytical/modeling tools and relevant assumptions, and eventually the decisions related to the representation of the risk maps to the end users (205, 206). In relation to the data aggregated by administrative divisions, commonly discussed issues include "edge effect" i.e., problems posed by the presence of adjacent locations not included in the analysis but that can influence its outcome, such as an unknown disease status in a country adjacent to the study area [(207, 208)]; and the "modifiable areal unit problem (MAUP)" i.e., the existence of differences in the analytical results obtained through the analysis of the same input data after aggregation at different levels. Examples include aggregation of point data from dairy farms in to counties or data available at sub districts level into provinces. The MAUP pertains to scale and zoning effect of the divisions (209, 210). A variety of methods are discussed in the literature to quantify and account for the edge effect and MAUP issues (211, 212). When spatial analytics and models are conducted based on available and potentially biased data, the resulting risk maps are invariably subjected to the negative impact of the data quality. However, we emphasize the use of existing data, bringing several databases together, and the spatiotemporal analytical tools can support initiating the process of improving data quality.

The choice of SAT, as discussed, varies with multiple factors. Inevitably, all analytical tools and models involve certain assumptions on statistical properties of variables and often these assumptions are violated in natural environments. In other words, none of the SAT are perfect matches for any particular situation (158). For example, spatial continuity of risk is a common assumption in risk-mapping process while there can be natural (e.g., mountain range acting as a physical barrier) or infrastructural barriers (e.g., urban vs. rural neighborhoods) that violate the continuity assumption resulting in step changes of risk between adjacent areas (112). Therefore, clarity on the choice of SAT, underlying assumptions, and the seven factors/criteria is essential when choosing SAT to address eco-epidemiological problems.

FUTURE DIRECTIONS

Improving the quality of spatially explicit health and environmental data through systematic collection of high-resolution data and public participation GIS approaches such as "crowdsourcing" or "citizen science data" is increasingly popular in both public and environmental health monitoring efforts (213–215). Additionally, the use of existing databases as passive surveillance systems and improving systematic data collection are suggested as ways to generate spatially explicit animal health databases (203).

While the geostatistical techniques introduced here, especially those under T4, commonly are frequentist approaches. The hierarchical specification of geostatistical models (216), therefore the adoption of a Bayesian framework for inference and suitable Gibbs sampling, MCMC, or INLA [(88); www.r-inla.org; (199)] for model fitting is being increasingly used. In addition to the geostatistical SATs discussed here, there are non-geostatistical spatial analytical tools such as Agent-based modeling (217–219) that are increasingly used by the researchers interested in spatial eco-epidemiological studies.

When modeling complex systems of adverse health and environmental effects, incorporation of several other analytical and modeling techniques in addition to SATs may support further exploring the phenomena including understanding the network effects (21). Spatial networks are another branch of the complex system approaches to spatial data. Because complex systems are often organized under the form of networks where nodes and edges are embedded in space, such as transportation networks of swine farms or water connectivity networks between salmon farms, the importance of connectivity in addition to the spatial proximity has a major role when determining disease transmission (220).

Predicting where the phenomenon would move/flow/spread next is an essential component in spatial modeling. SATs such as space-time kriging (T3 of **Table 1**) are capable of estimating such phenomena (221). Atmospheric dispersion models such as plume models (222) and Hybrid Single Particle Lagrangian Integrated Trajectory Model (HYSPLIT) (223) are examples of applications of spatial models that account for flow directions and cost surfaces used to predict wind-mediated transmission of arthropod-borne diseases. While these models can be considered as advanced spatiotemporal variations of SATs listed under T4 here, they can be computationally costly. Hence, for the researchers who are new to population-level spatial analysis and models, it is recommendable to start with the simpler and more established SATs to explore health or environmental threats prior to applying novel modeling techniques.

AUTHOR CONTRIBUTIONS

KK designed the framework, directed the review process, and wrote the article. JA and LZ provided expertise in methods and edited and reviewed the manuscript. AP contributed in design, expertise in methods, supervision, and revision of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded in part by the Minnesota Discovery, Research, and Innovation Economy (MnDRIVE) program and Office of the Vice President for Research (OVPR) of the University of Minnesota.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.00339/full#supplementary-material

REFERENCES

1. Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. *Environ Health Pers*. (2004) 112:998–1006. doi: 10.1289/ehp.6735
2. Mayer JD. The role of spatial analysis and geographic data in the detection of disease causation. *Pergamon Press Ltd Soc Sci Med*. (1983) 17:1213–21. doi: 10.1016/0277-9536(83)90014-X
3. March D, Susser E. The eco- in eco-epidemiology. *Int. J. Epidemiol*. (2006) 35:1379–83. doi: 10.1093/ije/dyl249
4. Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet*. (2012) 13:97–109. doi: 10.1038/nrg3142
5. Sanson RL, Pfeiffer DU, Morris RS. Geographic information systems: their application in animal disease control. *Rev Sci Tech*. (1991) 10:179–95 doi: 10.20506/rst.10.1.541
6. Ward MP, Carpenter TE. Analysis of time-space clustering in veterinary epidemiology. *Prev Vet Med*. (2000) 43:225–37. doi: 10.1016/S0167-5877(99)00111-7
7. Beale L, Abellan JJ, Hodgson S, Jarup L. Methodologic issues and approaches to spatial epidemiology. *Environ Health Pers*. (2008) 116:1105–10. doi: 10.1289/ehp.10816
8. Clements ACA, Pfeiffer DU. Emerging viral zoonoses: frameworks for spatial and spatiotemporal risk assessment and resource planning. *Vet J*. (2009) 182:21–30. doi: 10.1016/j.tvjl.2008.05.010
9. Richardson DB, Volkow ND, Kwan M, Kaplan RM, Goodchild MF, Croyle, RT. Spatial turn in health research. *Science*. (2013) 339:1390–92. doi: 10.1126/science.1232257
10. Kirby RS, Delmelle E, Eberth JM. Advances in spatial epidemiology and geographic information systems. *Ann Epidemiol*. (2017) 27:1–9. doi: 10.1016/j.annepidem.2016.12.001
11. Rezaeian M, Dunn G, St Leger S, Appleby L. Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. *J Epidemiol Commun Health*. (2007) 61:98–102. doi: 10.1136/jech.2005.043117
12. Goovaerts P. Geostatistics: a common link between medical geography, mathematical geology, and medical geology. *J South Afr Inst Min Metall*. (2014) 114:605–12.
13. Snow J. *Report of the Cholera Outbreak in the Parish of St. James, Westminster, During the Autumn of 1854*. London: John Churchill (1855). p. 97–120.
14. Koch T. *Disease Maps: Epidemics on the Ground*. Chicago, IL: University of Chicago Press. (2011). doi: 10.7208/chicago/9780226449401.001.0001
15. Cromley EK, McLafferty SL. *GIS and Public Health*. 2nd ed. New York, NY: The Guilford Publications Inc. (2011).
16. Wagner RS, Bauer SE, Vena JE. Integration of Different Epidemiologic Perspectives Applications to Spatial Epidemiology. In: Lawson AB. Banerjee S, Haining RP, Ugarte M, Chapman D, editors. *Handbook of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Boca Raton, FL: Hall/CRC Press Taylor Francis Group (2016). p. 3–2.
17. Stevens KB, Pfeiffer DU. The role of spatial analysis in risk-based animal disease management. In: Lawson AB, Banerjee S, Haining RP, Ugarte, M, Chapman D, editors. *Handbook of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Boca Raton, FL: Hall/CRC Press Taylor Francis Group (2016). p. 450–63.
18. Lawson AB. *Statistical Methods in Spatial Epidemiology*. 2nd Ed. Chichester: John Wiley & Sons Ltd. (2006) 3–20. doi: 10.1002/97804700 35771
19. Banerjee S, Carlin BP, Gelfand, AE. *Hierarchical Modeling and Analysis for Spatial Data, Second Edition, Monographs on Statistics and Applied Probability*. 2nd ed. Boca Raton, FL: Chapman and Hall/CRC, CRC Press, Taylor and Francis Group (2014).
20. Arbia G. *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*. Berlin: Springer-Verlag (2006).
21. Anselin L. Thirty years of spatial econometrics. *Papers Reg Sci*. (2010) 89:3–25. doi: 10.1111/j.1435-5957.2010.00279.x
22. Fortin MJ, James PMA, MacKenzie A, Melles SJ, Rayfield B. Spatial statistics, spatial regression, and graph theory in ecology. *Spat Stat*. (2012) 1:100–9. doi: 10.1016/j.spasta.2012.02.004
23. Carroll LN, Au AP, Detwiler LT, Fu TC, Painter IS, Abernethy NF. Visualization and analytics tools for infectious disease epidemiology: a systematic review. *J Biomed Inform*. (2014) 51:287–98. doi: 10.1016/j.jbi.2014.04.006
24. Cressie, NAC. *Statistics for Spatial Data. Wiley Series in Probability and Mathematical Statistics*. 2nd Revised ed. New York, NY: Wiley (1993) doi: 10.1002/9781119115151
25. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol*. (2005) 25:1965–78. doi: 10.1002/joc.1276
26. Fick SE, Hijmans, RJ. WorldClim 2: new 1-km spatial resolution climate surfacesfor global land areas. *Int J Climatol*. (2017) 37:4302–15. doi: 10.1002/joc.5086
27. Bright EA, Rose AN, Urban ML. LandScan 2012^{TM}. *High Resolution Global Population Data*. UT-Battelle, LLC. Oak Ridge National Laboratory. U.S. Department of Energy (2013). Available online at: https://landscan.ornl.gov/landscan-datasets (accessed July 06, 2020).
28. Perez AM, Thurmond MC, Grant PW, Carpenter TE. Use of the scan statistic on disaggregated province-based data: foot-and-mouth disease in Iran. *Prev Vet Med*. (2005) 71:197–207. doi: 10.1016/j.prevetmed.2005.07.005
29. Zandbergen PA. A comparison of address point, parcel and street geocoding techniques. *Comput Environ Urban Syst*. (2008) 32:214–32. doi: 10.1016/j.compenvurbsys.2007.11.006
30. Hart TC, Zandbergen PA. Reference data and geocoding quality examining completeness and positional accuracy of street geocoded crime incidents. *Policing Int J Police Strat Manag*. (2013) 36:263–94. doi: 10.1108/13639511311329705
31. Wieczorek AF, Delmerico AM. Geographic information systems. *Comput Stat*. (2009) 1:167–86. doi: 10.1002/wics.21
32. Isaaks EH, Srivastava, RM. *Applied Geostatistics*. New York, NY: Oxford University Press (1989).
33. Worton BJ. A convex hull-based estimator of home-range size. *Biometrics*. (1995) 51:1206–15. doi: 10.2307/2533254
34. Levefer DW. Measuring geographic concentration by means of the standard deviation ellipse. *Am J Sociol*. (1926) 32:88–94. doi: 10.1086/214027
35. Anselin L. The moran scatterplot as an ESDA tool to assess local instability in spatial association. In: Fischer M, Scholten H, Unwin D, editors. *Spatial Analytical Perspective on GIS*. London: Taylor and Francis (1996). p. 111–25. doi: 10.1201/9780203739051-8
36. Moore DA, Carpenter TE. Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiol Rev*. (1999) 21:143–6. doi: 10.1093/oxfordjournals.epirev.a017993
37. Moran PAP. Notes on continuous stochastic phenomena. *Biometrika*. (1950) 37:17–23. doi: 10.1093/biomet/37.1-2.17
38. Geary RC. The contiguity ratio and statistical mapping. *Incorporat Stat*. (1954) 5:115–45. doi: 10.2307/2986645
39. Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res*. (1967) 27:209–20.
40. Getis A, Ord JK. The analysis of spatial association by distance statistics. *Geogr Anal*. (1992) 24:189–206. doi: 10.1111/j.1538-4632.1992.tb00261.x
41. Ord JK, Getis A. Local spatial autocorrelation statistics: distributional issues and an application. *Geogr Anal*. (1995) 27:286–306. doi: 10.1111/j.1538-4632.1995.tb00912.x
42. Cliff A, Ord JK. Testing for spatial autocorrelation among regression residuals. *Geogr Anal*. (1972) 4:267–84. doi: 10.1111/j.1538-4632.1972.tb00475.x
43. Burridge P. On the cliff-ord test for spatial autocorrelation. *J R Stat Soc B*. (1980) 42:107–8. doi: 10.1111/j.2517-6161.1980.tb01108.x
44. Kelejian H, Robinson D. Spatial autocorrelation: a new computationally simple test with an application to per capita country police expenditures. *Reg Sci Urban Econ*. 22:317–31. doi: 10.1016/0166-0462(92)90032-V
45. Anselin L, Bera AK. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah A, Giles DE, editors. *Handbook of Applied Economic Statistics*. Marcel Dekker (1998). p. 237–89.
46. Clark PJ, Evans FC. Distance to nearest neighbor as a measure of spatial replationships in populations. *Ecology*. (1954) 35:445–53. doi: 10.2307/1931034
47. Ripley BD. Modelling spatial patterns. *J. R. Stat. Soc. B Stat Methodol*. (1977) 39:172–92. doi: 10.1111/j.2517-6161.1977.tb01615.x
48. Dixon PM. Ripley's K function. *Encyclopedia Environ*. (2002) 3:1796–803.

49. MacQueen JB. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Symposium on Math, Statistics, and Probability*. Berkeley, CA: University of California Press (1967). p. 281–97.
50. Anselin L. Local indicators of spatial association – LISA. *Geogr Anal*. (1995) 27:93–115. doi: 10.1111/j.1538-4632.1995.tb00338.x
51. Chen X, Wang K. Geographic area-based rate as a novel indicator to enhance research andprecision intervention for more effective HIV/AIDS control. *Prev Med Rep*. (2017) 5:301–7. doi: 10.1016/j.pmedr.2017.01.009
52. Oden N. Adjusting Moran's I for population density. *Stat Med*. (1995) 14:17–26. doi: 10.1002/sim.4780140104
53. Jacquez GM, Estberg L, Greiling D, Durbeck H, Do E, Long A, Rommel B. ClusterSeer v.2.05. In: *User Manual book 2. Software for the Detection and Analysis of Event Clusters*. Ann Arbor, MI: BioMedware Inc. (2012)
54. Fotheringham AS. Trends in quantitative methods I: stressing the local. *Progr Hum Geogr*. (1997) 21:88–96. doi: 10.1191/030913297676693207
55. Fotheringham AS. Trends in quantitative methods III: stressing the visual. *Progr Hum Geogr*. (1999) 23:597–606 doi: 10.1191/0309132996677 56016
56. Cuzick J, Edwards R. Spatial clustering for inhomogenous populations. *J R Stat Soc B*. (1990) 52:73–104. doi: 10.1111/j.2517-6161.1990.tb01773.x
57. Kulldorff M. A spatial scan statistic. *Commun Stat Theory Methods*. (1997) 26:1481–96. doi: 10.1080/03610929708831995
58. Kulldorff M, Nagarwalla N. Spatial disease clusters and inference. *Stat Med*. (1995) 14:799–810. doi: 10.1002/sim.4780140809
59. Kulldorff M. *SaTScanTMv9.4.1: Software for the Spatial and Space-Time Scan Statistics*. (2009). Available online at: http://www.satscan.org/ (accessed July 06, 2020).
60. Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *Int J Health Geogr*. (2005) 4:11. doi: 10.1186/1476-072X-4-11
61. Turnbull BW, Iwano EJ, Burnett W, Howe HL, Clark LC. Monitoring for clusters of disease: applications to leuke mia incidence in upstate New York. *Am J Epidemiol*. (1990) 132:S136–43. doi: 10.1093/oxfordjournals.aje.a115775
62. Besag J, Newell J. The detection of clusters in rare diseases. *J R Stat Soc A*. (1991) 154:143–55. doi: 10.2307/2982708
63. Knox EG. The detection of space-time interactions. *J R Stat Soc C Appl Stat*. (1964) 13:25–30. doi: 10.2307/2985220
64. Barton DE, David FN, Merrington M. A criterion for testing contagion in time and space. *Ann Hum Genet*. (1965) 29:97–102. doi: 10.1111/j.1469-1809.1965.tb00504.x
65. Jacquez GM. A k nearest neighbour test for space-time interaction. *Stat Med*. (1996) 15:1935–1949. doi: 10.1002/(SICI)1097-0258(19960930)15:183.0.CO;2-I
66. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. *PLoS Med*. (2005) 2:216–24. doi: 10.1371/journal.pmed.0020059
67. Kulldorff M, Glaz J, Pozdnyakov V, Wallenstein S, Marcelo Azevedo C. Applications of spatial scan statistics: a review. In: Glaz J, Vladmir P, Sylvan W, editors. *Scan Statistics: Methods and Applications*. Birkhauser (2009). p. 129–152. doi: 10.1007/978-0-8176-4749-0
68. Ederer F, Myers MH, Mantel N. A statistical problem in space and time: do leukemia cases come in clusters? *Biometrics*. (1964) 20:626–38. doi: 10.2307/2528500
69. Stark CR, Mantel N. Lack of seasonal or temporal spatial clustering of Down's Syndrome births inMichigan. *Am J Epidemiology*. (1967) 86:199–213. doi: 10.1093/oxfordjournals.aje.a120725
70. Jacquez GM, Greiling D, Durbeck H, Estberg L, Do E, Long E, et al. ClusterSeer. In: *User Guide 2: Software for Identifying Disease Clusters*. Ann Arbor, MI: TerraSeer Press (2002). p. 316.
71. Bilonick RA. Monthly hydrogen ion deposition maps for the northeastern U.S. from July 1982 to September 1984. *Atmos Environ*. (1988) 22:1909–24. doi: 10.1016/0004-6981(88)90080-7
72. Snepvangers JJJC, Heuvelink GBM, Huisman JA. Soilwater content interpolation usingspatio-temporal kriging with external drift. *Geoderma*. (2003). 112:253–71. doi: 10.1016/S0016-7061(02)00310-5
73. Bithell J. An application of density estimation to geographical epidemiology. *Stat Med*. (1990) 9:691–701. doi: 10.1002/sim.4780090616
74. Kelsall JE, Diggle PJ. Kernel estimation of relative risk. *Bernoulli*. (1995) 1:3–16. doi: 10.2307/3318678
75. Hazelton ML. Kernel smoothing methods. In: Lawson AB, Banerjee S, Haining RP, Ugarte MD, editors. *Handbooks of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Boca Raton, FL: Chapman and Hall/CRC Press (2016). p. 195–205.
76. Silverman BW. *Density Estimation for Statistics and Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC (1986).
77. Wand MP, Jones MC. *Kernel Smoothing*. Boca Raton, FL: Chapman Hall/CRC. (1995). doi: 10.1007/978-1-4899-4493-1
78. Cressie N, Read TRC. Spatial data analysis of regional counts. *Biometrical J*. (1989) 31:699–719. doi: 10.1002/bimj.4710310607
79. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*. (1987) 43:671–81. doi: 10.2307/2532003
80. Aurenhammer F. Voronoi diagrams – a survey of a fundamental geometric data structure. *ACM Comput Surveys*. (1991) 23:345–405. doi: 10.1145/116873.116880
81. Tomczak M. Spatial interpolation and its uncertainty using automated anisotropic inverse distance weighting (IDW) - cross-validation/jackknife approach. *J Geogr Inform Decis Anal*. (1998) 2:18–30.
82. Lu GY, Wong DW. An adaptive inverse-distance weighting spatial interpolation technique. *Comput Geosci*. (2008) 34:1044–55. doi: 10.1016/j.cageo.2007.07.010
83. Burrough PA, McDonnell RA. *Principles of Geographical Information Systems*. Oxford: Oxford University Press (1998).
84. Li J, Heap AD. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecol Inform*. (2011) 6:228–41. doi: 10.1016/j.ecoinf.2010.12.003
85. Kafidar K. Smoothing geographical data, particularly rates of disease. *Stat Med*. (1996) 15:2539–60. doi: 10.1002/(SICI)1097-0258(19961215)15:23<2539::AID-SIM379>3.0.CO;2-B
86. Mungiole M, Pickle LW, Simonson KH. Application of a weighted headbanging algorithm to mortality data maps. *Stat Med*. (1999) 18:3201–9. doi: 10.1002/(sici)1097-0258(19991215)18:23<3201::aid-sim310>3.0.co;2-u
87. Gelman A, Price PN, Lin C. A method for quantifying artefacts in mapping methods illustrated by application to headbanging. *Stat Med*. (2000) 19:2309–20. doi: 10.1002/1097-0258(20000915/30)19:17/18<2309::aid-sim571>3.0.co;2-h
88. Rue H, Martino S, Chopin N. Approximate bayesian inference for latent Gaussian models using integrated nested laplace approximations (with discussion). *J R Stat Soc B*. (2009) 71:319–392. doi: 10.1111/j.1467-9868.2008.00700.x
89. Lindgren F, Rue H, Lindstrom J. An explicit link between gaussian fields and Gaussian Markov random fields: the SPDE approach (with discussion). *J R Stat Soc B*. (2011) 73:423–98. doi: 10.1111/j.1467-9868.2011.00777.x
90. Matheron G. *The Theory of the Regionalized Variables and its Applications*. Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau No. 5. Paris: École Nationale Supérieure des Mines de Paris (1971). p. 211.
91. Wackernagel H. *Multivariate Geostatistics: An Introduction with Applications*. 3rd ed. Berlin; Heidelberg: Springer-Verlag (2003). doi: 10.1007/978-3-662-05294-5_1
92. MacNab YC, Gustafson P. Regression B-spline smoothing in bayesian disease mapping: with an application to patient safety surveillance. *Stat Med*. (2007) 26:4455–74. doi: 10.1002/sim.2868
93. Lusting LK. Trend-surface analysis of the Basin Range Province, some geomorphic implications. *U.S. Geological Survey Professional Paper 500-D*. Washington DC: U.S. Government printing office (1969). p. 70. doi: 10.3133/pp500D
94. Davis JC. *Statistics and Data Analysis in Geology*. New York, NY: John Wiley and Sons (1973). p. 550.
95. Agterberg FP. Trend surface analysis. In: Gaile GL, Willmott CJ, editors. *Spatial Statistics Models*. Dordrecht: Riedel (1984). p. 147–71. doi: 10.1007/978-94-017-3048-8_8

96. Mitas L, Mitasova H. Spatial Interpolation. In:.Longley P, Goodchild MF, Maguire DJ, Rhind DW, editors. *Geographical Information Systems: Principles, Techniques, Management and Applications*. Hoboken, NJ: Wiley (1999), p.481–92.
97. Stein A, Corsten LCA. Universal kriging and co-kriging as a regression procedure. *Biometrics*. (1991) 47:575–87. doi: 10.2307/2532147
98. Odeh IOA, McBratney AB, Chittleborough DJ. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma*. (1995) 67:215–26. doi: 10.1016/0016-7061(95)00007-B
99. Goovaerts P. *Geostatistics for Natural Resource Evaluation*. *Applied Geostatistics*. New York, NY: Oxford University Press (1997).
100. Eldeiry AA, Garcia LA. Comparison of ordinary kriging, regression kriging, and cokriging techniques to estimate soil salinity using LANDSAT images. *J Irrig Drain Eng*. (2010) 136:355–64. doi: 10.1061/(ASCE)IR.1943-4774.0000208
101. Cressie N, Wikle CK. *Statistics for Spatio-Temporal Data*. *Wiley Series in Probability and Statistics* John Wiley & sons Inc. (2011).
102. Biondi F. Space-time kriging extension of precipitation variability at 12 km spacing from tree-ring chronologies and its implications for drought analysis. *Hydrol Earth Syst Sci Discuss*. (2013) 10:4301–35. doi: 10.5194/hessd-10-4301-2013
103. MacNab YC, Dean CB. Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics*. (2004) 57:949–56. doi: 10.1111/j.0006-341X.2001.00949.x
104. Anselin L. *Spatial Econometrics: Methods and Models*. Boston, MA; Dordrecht: Kluwer Academic Publishers (1988).
105. Anselin L. Spatial Econometrics. In: Baltagi BH editor. *A Companion to Theoretical Econometrics*. Blackwell Publishing Ltd. (2001) 310–30. doi: 10.1002/9780470996249.ch15
106. Haining R. *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press (2003). doi: 10.1017/CBO9780511754944
107. Brunsdon C, Fotheringham AS, Charlton ME. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geogr Anal*. (1996) 28:281–98. doi: 10.1111/j.1538-4632.1996.tb00936.x
108. Nakaya T, Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted poisson regression for disease association mapping. *Stat Med*. (2005) 24:2695–717. doi: 10.1002/sim.2129
109. Besag J, York J, Molli'e A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math*. (1991) 43:1–20. doi: 10.1007/BF00116466
110. Leroux BG, Lei X, Breslow N. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: Miller Jr. W, Editor. *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. Springer. (2000). p. 179–91. doi: 10.1007/978-1-4612-1284-3_4
111. Mariella L, Tarantino M. Spatial temporal conditional auto-regressive model: a new autoregressive matrix. *Austr J Stat*. (2010) 39:223–44. doi: 10.17713/ajs.v39i3.246
112. Rushworth A, Lee D, Sarran C. An adaptive spatiotemporal smoothing model for estimating trends and step changes in disease risk. *J R Stat Soc C Appl Stat*. (2017) 66:141–57. doi: 10.1111/rssc.12155
113. Lawson AB, Choi J, Cai B, Hossain M, Kirby RS, Liu JH. Bayesian 2-stage space-time mixture modeling with spatial misalignment of the exposure in small area health data. *J Agric Biol Environ Stat*. (2012) 17:417–41. doi: 10.1007/s13253-012-0100-3
114. Hossain MM, Lawson AB. Space-time bayesian small area disease risk models: development and evaluation with a focus on cluster detection. *Environ Ecol Stat*. (2010) 17:73–95. doi: 10.1007/s10651-008-0102-z
115. Cai B, Lawson AB, Hossain MM, Choi J. Bayesian latent structure models with space-time dependent covariates. *Stat Modell*. (2012) 12:145–64. doi: 10.1177/1471082X1001200202
116. Berman M, Turner R. Approximating point process likelihoods with GLIM. *Appl Stat*. (1992) 41:31–8. doi: 10.2307/2347614
117. Hossain MM, Lawson AB. Approximate methods in bayesian point process spatial models. *Comput Stat Data Anal*. (2009) 53:2831–42. doi: 10.1016/j.csda.2008.05.017
118. Diggle P, Rowlingson B. A conditional approach to point process modeling of elevated risk. *J Royal Stat Soc Series A*. (1994) 157:433–40. doi: 10.2307/2983529
119. Diggle P. Spatio-temporal point processes. In: Finkenstadt B, Held L, Isham V, editors. *Methods Applications. Statistical Methods for Spatio-Temporal Systems* London: CRC Press (2007). p. 1–45. doi: 10.1201/9781420011050.ch1
120. Nijkamp P. Reflections on gravity and entropy models. *Reg Sci Urban Econ*. (1975) 5:203–25. doi: 10.1016/0166-0462(75)90004-6
121. Bailey T, Gatrell A. *Interactive Spatial Data Analysis*. Essex: Longman (1996). doi: 10.1016/0277-9536(95)00183-2
122. Barrios JM, Verstraeten WW, Maes P, Aerts JM, Farifteh J, Coppin P. Using the gravity model to estimate the spatial spread of vector-borne diseases. *Int J Environ Res Public Health*. (2012) 9:4346–64. doi: 10.3390/ijerph9124346
123. Truscott J, Ferguson NM. Evaluating the adequacy of gravity models as a description of human mobility for epidemic modelling. *PLoS Comput Biol*. (2012) 8:e1002699. doi: 10.1371/journal.pcbi.1002699
124. Banerjee S. Spatial survival models. In: Lawson AB, Banerjee S, Haining RP, Ugarte MD, editors. *Handbooks of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Boca Raton, FL: Chapman Hall/CRC Press Taylor Francis Group (2016). p. 303–311.
125. Elith J, Phillips SJ, Hastie T, Dudik M, Chee YE, Yates CJ. A statistical explanation of MaxEnt for ecologists. *Divers Distrib*. (2011) 17:43–57. doi: 10.1111/j.1472-4642.2010.00725.x
126. Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecol Model*. (2006) 190:231–59. doi: 10.1016/j.ecolmodel.2005.03.026
127. Merow C, Smith MJ, Silander JA Jr. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*. (2013) 36:001–12. doi: 10.1111/j.1600-0587.2013.07872.x
128. Stockwell DRB, Peters DG. The GARP modelling system: problems and solutions to automated spatial prediction. *Int J Geogr Inform Syst*. (1999) 13:143–58. doi: 10.1080/136588199241391
129. Stockwell DRB, Peterson AT. Effects of sample size on accuracy of species distribution models. *Ecol Model*. (2002) 148:1–13. doi: 10.1016/S0304-3800(01)00388-X
130. Blackburn JK, McNyset KM, Curtis A, Hugh-Jones ME. Modeling the geographic distribution of bacillus anthracis, the causative agent of anthrax disease, for the contiguous United States using predictive ecologic niche modeling. *Am J Trop Med Hygiene*. (2007) 77:1103–10. doi: 10.4269/ajtmh.2007.77.1103
131. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324
132. Mi C, Huettmann F, Guo Y, Han X, Wen L. Why choose random forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ*. (2016) 5:e2849. doi: 10.7717/peerj.2849
133. Hastie T, Tibshirani R. Generalized Additive Models. London: Chapman and Hall (1990).
134. Leathwick JR, Elith J, Hastie T. Comparative performance of generalised additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecol Model*. (2006) 199:188–96. doi: 10.1016/j.ecolmodel.2006.05.022
135. Wood SN. *Generalized Additive Models: An Introduction with R*. Boca Raton FL: Chapman & Hall/CRC (2006) doi: 10.1201/9781420010404
136. Civco DL. Artificial neural networks for land-cover classification and mapping. *Int J Geogr Inform Sci*. (1993) 7:173–86. doi: 10.1080/02693799308901949
137. Sordo M. Introduction to neural networks in healthcare. *Open Clin*. (2002). p. 1–7. Available online at: http://www.openclinical.org/docs/int/neuralnetworks011.pdf (accessed on April 14, 2020).
138. Li J. A critical review of spatial predictive modeling process in environmental sciences with reproducible examples in R. *Appl Sci*. (2019) 9:2048. doi: 10.3390/app9102048
139. Maliene V, Grigonis V, Palevicius V, Griffiths S. Geographic information system: old principles with new capabilities. *Urban Des Int*. (2011) 16:1–6. doi: 10.1057/udi.2010.25

140. Tobler WR. A computer movie simulating urban growth in the detroit region. *Econ Geogr*. (1970) 46:234–40. doi: 10.2307/143141
141. Dormann CF, McPherson JM, Araujo MB, Bivand R, Bolliger J, Carl G, et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*. (2007) 30:609–28. doi: 10.1111/j.2007.0906-7590.05171.x
142. Salima BA, Bellefon MD. Spatial autocorrelation indices. In: Loonis V, Bellefon MD, editors. *Handbook of Spatial Analysis: Theory Aplication with R*. (2018). p. 51–68.
143. Jacquez GM. Spatial cluster analysis: In: Fotheringham S, Wilson J, editors. *The Handbook of Geographic Information Science*. Blackwell Publishing (2008). p. 395–416. doi: 10.1002/9780470690819.ch22
144. Costa MA, Assuncao RM. A fair comparison between the spatial scan and the besag-newell disease clustering tests. *Environ Ecol Stat*. (2005) 12:301–19. doi: 10.1007/s10651-005-1515-6
145. Patil GP, Taillie C. Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environ Ecol Stat*. (2004) 11:183–97. doi: 10.1023/B:EEST.0000027208.48919.7e
146. Jacquez GM, Kaufmann A, Goovaerts P. Boundaries, links and clusters: a new paradigmin spatial analysis? *Environ Ecol Stat*. (2008) 15:403–19. doi: 10.1007/s10651-007-0066-4
147. Waller LA, Turnbull BW. The Effects of scale on tests for disease clustering. *Stat Med*. (1993) 12:1869–84. doi: 10.1002/sim.4780121913
148. Morris RD, Munasinghe RL. Aggregation of existing geographic regions to diminish spurious variability of disease rates. *Stat Med*. (1993) 12:1915–29. doi: 10.1002/sim.4780121916
149. Berke O. Exploratory spatial relative risk mapping. *Prev Vet Med*. (2005) 71:173–82. doi: 10.1016/j.prevetmed.2005.07.003
150. Diggle P, Rowlingson B, Su T. Point process methodology for on-line spatiotemporal disease surveillance. *Environmetrics*. (2005) 16:423–34. doi: 10.1002/env.712
151. Brunsdon C, Fotheringham S, Charlton M. Geographically weighted summary statistics: a framework for localised exploratory data analysis. *Comput Environ Urban Syst*. (2002) 26:501–24. doi: 10.1016/S0198-9715(01)00009-6
152. Rushton G. Improving the geographic basis of health surveillance using GIS. In: Gatrell AC, Loytonen M, editors. *GIS and Health*. London: Taylor and Francis (1998). p. 63–79.
153. Rushton G, Peleg I, Banerjee A, Smith G, West M. Analyzing geographic patterns of disease incidence: rates of late-stage colorectal cancer in Iowa. *J Med Syst*. (2004) 28:223–36. doi: 10.1023/b:joms.0000032841.39701.36
154. Rogers DJ, Sedda L. Statistical models for spatially explicit biological data. *Parasitol*. (2012) 139:1852–69. doi: 10.1017/S0031182012001345
155. Kanankege KST, Alkhamis MA, Phelps NBD, Perez AM. A probability co-kriging model to account for reporting bias and recognize areas at high risk for zebra mussels and Eurasian watermilfoil invasions in Minnesota. *Front Vet Sci*. (2018) 4:231. doi: 10.3389/fvets.2017.00231
156. Huanxin W, Nan H. Trend-surface analysis was applied to the study of the geographic distribution for LBWI in China. *Chinese J Hlth Stat*. (1991) 8:12.
157. Watson GS. *Trend Surface Analysis and Spatial Correlation*. Technical Report 124. AD 699 163. Department of Statistics, The John Hopkins University, Baltimore, MD (1969). Available online at: https://apps.dtic.mil/dtic/tr/fulltext/u2/699163.pdf
158. Susanto F, de Souza P, He J. Spatiotemporal interpolation for environmental modelling. *Sensors*. (2016) 16:1245. doi: 10.3390/s16081245
159. Frome EL, Checkoway H. Use of poisson regression models in estimating incidence rates and ratios. *Am J Epidemiol*. (1985) 121:309–23. doi: 10.1093/oxfordjournals.aje.a114001
160. Anselin L. Under the hood: issues in the specification and interpretation of spatial regression models. *Agric Econ*. (2002) 27:247–67. doi: 10.1111/j.1574-0862.2002.tb00120.x
161. Auchincloss AH, Gebreab SY, Mair C, Roux AVD. A review of spatial methods in epidemiology, 2000-2010. *Annu Rev Public Health*. (2012) 33:107–22. doi: 10.1146/annurev-publhealth-031811-124655
162. Wall MM. A close look at the spatial structure implied by the CAR and SAR models. *J Stat Plan Inference*. (2004) 121:311–24. doi: 10.1016/S0378-3758(03)00111-3
163. Fotheringham AS, Charlton ME, Brunsdon C. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environ Plan A*. (1998) 30:1905–27. doi: 10.1068/a301905
164. Lawson AB. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. Boca Raton, FL: Chapman and Hall/CRC Press; Taylor & Francis Group (2013).
165. Lee D, Rushworth A, Napier G. CARBayesST: spatio-temporal areal unit modeling in R with conditional autoregressive priors using the CARBayesST package. *J Stat Softw*. (2018) 84, 1–39. doi: 10.18637/jss.v084.i09
166. Paciorek CJ. Computational techniques for spatial logistic regression with large datasets. *Comput Stat Data Anal*. (2007) 51:3631–53. doi: 10.1016/j.csda.2006.11.008
167. Peterson AT, Sanchez-Cordero V, Ben Beard C, Ramsey JM. Ecologic niche modeling and potential reservoirs for chagas disease, Mexico. *Emerg Infect Dis*. (2002) 8:662–7. doi: 10.3201/eid0807.010454
168. Escobar LE, Craft ME. Advances and limitations of disease biogeography using ecological niche modeling. *Front Microbiol*. (2016) 7:1174. doi: 10.3389/fmicb.2016.01174
169. Peterson AT, Tian H, Martínez-Meyer E, Soberón J, Sánchez-Cordero V, Huntley B. Modeling distributional shifts of individual species biomes. In: Lovejoy TE, Hannah L, editors. *Climate Change and Biodiversity*. New Haven, CT: Yale University Press (2005). p. 211–28.
170. Fielding AL, Bell JF. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Envron Conserv*. (1997) 24:38–49. doi: 10.1017/S0376892997000088
171. Lusted LB. ROC recollected. *Editorial Med Decisi Making*. (1984) 4:131–4. doi: 10.1177/0272989X8400400201
172. Akaike H. On entropy maximization principle. In: Krishnaiah PR editor. *Applications of Statistics*. Amsterdam: North-Holland Publishing Co. (1977). p. 27–41.
173. Schwarz GE. Estimating the dimension of a model. *Ann Stat*. (1978) 6:461–4. doi: 10.1214/aos/1176344136
174. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measuresof model complexity and fit (with discussion). *J R Stat Soc B*. (2002) 64:583–639. doi: 10.1111/1467-9868.00353
175. van der Linde A. (2005). DIC in variable selection.Statistica. *Neerlandica*. 1:45–56. doi: 10.1111/j.1467-9574.2005.00278.x
176. Watanabe S. Asymptotic equivalence of bayes crossvalidation and widely applicableinformation criterion in singular learning theory. *J Mach Learn Res*. (2010) 11:3571–94. Available online at: http://www.jmlr.org/papers/volume11/watanabe10a/watanabe10a.pdf
177. Lee H, Ghosh S. Performance of information criteria for spatial models. *J Stat Comput Simul*. (2009) 79:93–106. doi: 10.1080/00949650701611143
178. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. *Stat Comput*. (2014) 24:997–1016. doi: 10.1007/s11222-013-9416-2
179. Anselin L. Model validation inspatial econometrics: a review and evaluation of alternative approaches. *Int Reg Sci Rev*. (1988) 11:279–316. doi: 10.1177/016001768801100307
180. Buckland ST, Elston DA. Empirical models for the spatial distribution of wildlife. *J Appl Ecol*. (1993) 30:478–95. doi: 10.2307/2404188
181. Verbyla DL, Litaitis JA. Resampling methods for evaluating classification accuracy of wildlife habitat models. *Environ Manag*. (1989) 13:783–7. doi: 10.1007/BF01868317
182. Capen DE, Fenwick JW, Inkley DB, Boynton AC. Multivariate models of songbird habitat in New England forests. In: Verner JA, Morrison ML, Ralph CJ, editors. *Wildlife 2000: Modelling Habitat Relationships of Terrestrial Vertebrates*. Madison WI: University of Wisconsin Press (1986). p. 171–75.
183. Fielding AH, Haworth PF. Testing the generality of bird-habitat models. *Conserv Biol*. (1995) 9:1466–81. doi: 10.1046/j.1523-1739.1995.09061466.x
184. Stockwell DRB. *Machine learning and the problem of prediction and explaination in ecological modeling* (Ph.D. thesis). Australian National University, Canberra, ACT, Australia (1992).
185. Kohavi R. A study of cross-validation bootstrap for accuracy estimation andmodel selection. In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, Montreal QC (1995). p. 1137–43.
186. Anselin L. From spacestat to CyberGIS: twenty years of spatial data analysis software. *Int Reg Sci Rev*. (2012) 35:131–57. doi: 10.1177/0160017612438615

187. Kuhn W, Ballatore A. Designing a language for spatial computing. In: Bacao F, Santos MY, Painho M, editors. *AGILE*. Springer International Publishing (2015). p. 309–326. doi: 10.1007/978-3-319-16787-9_18
188. *ESRI ArcMap Version 10.7*. Redlands CA: Environmental Research Institute Inc. (2019).
189. QGIS Development Team. *QGIS Geographic Information System*. Open Source Geospatial Foundation (2018). Available online at: https://qgis.org (accessed July 06, 2020).
190. GRASS Development Team. *Geographic Resources Analysis Support System (GRASS) Software, Version 7.2*. Open Source Geospatial Foundation. Electronic Document (2017). Available online at: http://grass.osgeo.org
191. Anselin L, Syabri I, Kho Y. GeoDa: an introduction to spatial data analysis. *Geogr Anal*. (2006) 38:5–22. doi: 10.1111/j.0016-7363.2005.00671.x
192. Levine N. *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.3)*. Houston, TX; Washington, DC: Ned Levine Associates; the National Institute of Justice (2010).
193. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna (2017). Available online at: https://grass.osgeo.org (accessed July 06, 2020).
194. SAS Institute Inc. *SAS/STAT® 14.1*. Cary, NC: SAS Institute Inc. (2015).
195. StataCorp. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp, LLC. (2019).
196. Bivand RS, Hauke J, Kossowski T. Computing the jacobian in gaussian spatial autoregressive models: an illustrated comparison of available methods. *Geogr Anal*. (2013) 45:150–79. doi: 10.1111/gean.12008
197. Bivand R, Piras G. Comparing implementations of estimation methods for spatial econometrics. *J Stat Softw*. (2015) 63:1–36. doi: 10.18637/jss.v063.i18
198. Lee D. CARBayes: an R package for bayesian spatial modeling with conditional autoregressive priors. *J Stat Softw*. (2013) 55:1–24. doi: 10.18637/jss.v055.i13
199. Schrodle B, Held L. A primer on disease mapping and ecological regression using INLA. *Comput Stat*. (2011) 26:241–58. doi: 10.1007/s00180-010-0208-2
200. Kanankege KST. *The use of spatiotemporal analytical tools to inform decisions and policy in One Health scenarios* (Ph.D. dissertaion) (2019). Available online at: http://hdl.handle.net/11299/202418 (accessed July 06, 2020).
201. Sternbach G. The history of anthrax. *J Emerg Med*. (2003) 24:463–7. doi: 10.1016/S0736-4679(03)00079-9
202. Driks A. The bacillus anthracis spore. *Mol Aspects Med*. (2009) 30:368–73. doi: 10.1016/j.mam.2009.08.001
203. Kanankege KST, Abdrakhmanov SK, Alvarez J, Glaser L, Bender JB, Mukhanbetkaliyev YY, et al. Comparison of spatiotemporal patterns of historic animal Anthrax outbreaks in Minnesota and Kazakhstan. *PLoS ONE*. (2019) 14:e0217144. doi: 10.1371/journal.pone.0217144
204. Mullins JC, Garofolo G, Van Ert M, Fasanella A, Lukhnova L, Hugh-Jones ME, et al. Ecological niche modeling of *Bacillus anthracis* on three continents: evidence for genetic-ecological divergence? *Plos One*. (2013) 8:8. doi: 10.1371/journal.pone.0072451
205. Ocaña-Riola R. Common errors in disease mapping. *Geospat Health*. (2010) 4:139–54. doi: 10.4081/gh.2010.196
206. Loth L, Gilbert M, Wu J, Czarnecki C, Hidayat M, Xiao X. Identifying risk factors of highly pathogenic avian influenza (H5N1 Subtype) in Indonesia. *Prev Vet Med*. (2011) 102:50–8. doi: 10.1016/j.prevetmed.2011.06.006
207. Griffith DA, Amrhein CG. An evaluation of correction techniques for boundary effects in spatial statistical analysis: traditional methods. *Geogr Anal*. (1983) 15:352–60. doi: 10.1111/j.1538-4632.1983.tb00794.x
208. Griffith DA. An evaluation of correction techniques for boundary effects in spatial statistical analysis: contemporary methods. *Geogr Anal*. (1985) 17:81–8. doi: 10.1111/j.1538-4632.1985.tb00828.x
209. Openshaw S. The modifiable areal unit problem. In: *Concepts and Techniques in Modern Geography*. Norwich: Geo Books (1984). Available online at: https://www.uio.no/studier/emner/sv/iss/SGO9010/openshaw1983.pdf
210. Fotheringham AS, Wong DWS. The modifiable areal unit problem in multivariate statistical analysis. *Environ Plan A*. (1991) 23:1025–44. doi: 10.1068/a231025
211. Laurance WF, Nascimento HEM, Laurance SG, Andrade A, Ewers RM, Harms KE, et al. Habitat fragmentation, variable edge effects, and the landscape-divergence hypothesis. *Plos One*. (2007) 2:e1017. doi: 10.1371/journal.pone.0001017
212. Mu L, Wang F. A scale-space clustering method: mitigating the effect of scale in the analysis zone-based data. *Ann Assoc Am Geogr*. (2008) 98:85–101. doi: 10.1080/00045600701734224
213. Linard C, Tatem AJ. Large-scale spatial population databases in infectious disease research. *Int J Health Geogr*. (2012) 11:7. doi: 10.1186/1476-072X-11-7
214. Tatem AJ, Adamo S, Bharti N, Burgert CR, Castro M, Dorelien A, et al. Mapping populations at risk: improving spatial demographic data for infectious disease modeling and metric derivation. *Popul Health Metr*. (2012) 10:8. doi: 10.1186/1478-7954-10-8
215. Sun CC, Royle JA, Fuller AK. Incorporating citizen science data in spatially explicit integrated population models. *Ecology*. (2019) 100:e02777. doi: 10.1002/ecy.2777
216. Gelfand AE, Banerjee S. Bayesian modeling and analysis of geostatistical data. In: *Annual Review of Statistics and Its Application, Vol 4. Palo Alto: Annual Reviews* (2017) p. 245–66.
217. Auchincloss AH, Diez Roux AV. A new tool for epidemiology: the usefulness of dynamic-agent models in understanding place effects on health. *Am J Epidemiol*. (2008) 168:1–8. doi: 10.1093/aje/kwn118
218. Crooks AT, Heppenstall, AJ. Agent-based models of geographical systems. Heppenstall AJ, Crooks AT, See LM, Batty, M, editors. Heidelberg, NY: Springer Dordrecht. (2012) 219–52. doi: 10.1007/978-90-481-8927-4
219. Tracy M, Cerda M, Keyes KM. Agent-based modeling in public health: current applications and future directions. *Annu Rev Public Health*. (2018) 39:77–94. doi: 10.1146/annurev-publhealth-040617-014317
220. Craft ME. Infectious disease transmission and contact networks in wildlife and livestock. *Philos Trans R Soc B Biol Sci*. (2015) 370:20140107. doi: 10.1098/rstb.2014.0107
221. Iglesias I, Montes F, Martínez M, Perez A, Gogin A, Kolbasov D, et al. Spatio-temporal kriging analysis to identify the role of wild boar in the spread of African swine fever in the Russian federation. *Spat Stat*. (2018) 28:226–35. doi: 10.1016/j.spasta.2018.07.002
222. Gloster J, Mellor PS, Manning AJ, Webster HN, Hort MC. Assessing the risk of windborne spread of bluetongue in the 2006 outbreak of disease in northern Europe. *Vet Rec*. (2007) 160:54–56. doi: 10.1136/vr.160.2.54
223. Durr PA, Graham K, van Klinken RD. Sellers' revisited: a big data reassessment of historical outbreaks of bluetongue and African horse sickness due to the long-distance wind dispersion of culicoides midges. *Front Vet Sci*. (2017) 4:98. doi: 10.3389/fvets.2017.00098

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

REVIEW
published: 04 September 2020
doi: 10.3389/fvets.2020.00596

Multilayer and Multiplex Networks: An Introduction to Their Use in Veterinary Epidemiology

Amy C. Kinsley[1], Gianluigi Rossi[2], Matthew J. Silk[3,4] and Kimberly VanderWaal[1]*

[1] Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, St. Paul, MN, United States, [2] Roslin Institute and Royal (Dick) School of Veterinary Sciences, University of Edinburgh, Edinburgh, United Kingdom, [3] Centre for Ecology and Conservation, University of Exeter Penryn Campus, Penryn, United Kingdom, [4] Environment and Sustainability Institute, University of Exeter, Penryn, United Kingdom

Contact network analysis has become a vital tool for conceptualizing the spread of pathogens in animal populations and is particularly useful for understanding the implications of heterogeneity in contact patterns for transmission. However, the transmission of most pathogens cannot be simplified to a single mode of transmission and, thus, a single definition of contact. In addition, host-pathogen interactions occur in a community context, with many pathogens infecting multiple host species and most hosts being infected by multiple pathogens. Multilayer networks provide a formal framework for researching host-pathogen systems in which multiple types of transmission-relevant interactions, defined as network layers, can be analyzed jointly. Here, we provide an overview of multilayer network analysis and review applications of this novel method to epidemiological research questions. We then demonstrate the use of this technique to analyze heterogeneity in direct and indirect contact patterns amongst swine farms in the United States. When contact among nodes can be defined in multiple ways, a multilayer approach can advance our ability to use networks in epidemiological research by providing an improved approach for defining epidemiologically relevant groups of interacting nodes and changing the way we identify epidemiologically important individuals such as superspreaders.

Keywords: network analysis, multilayer networks, animal movement, pigs, transmission, infectious disease

OPEN ACCESS

Edited by:
Salome Dürr,
University of Bern, Switzerland

Reviewed by:
Vitaly Belik,
Freie Universität Berlin, Germany
Krishna Thakur,
University of Prince Edward Island, Canada
Guillaume Fournié,
Royal Veterinary College (RVC), United Kingdom

****Correspondence:***
Amy C. Kinsley
carr0603@umn.edu

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *03 December 2019*
Accepted: *27 July 2020*
Published: *04 September 2020*

Citation:
Kinsley AC, Rossi G, Silk MJ and VanderWaal K (2020) Multilayer and Multiplex Networks: An Introduction to Their Use in Veterinary Epidemiology. Front. Vet. Sci. 7:596. doi: 10.3389/fvets.2020.00596

INTRODUCTION

The use of social network analysis and modeling in epidemiology has significantly enhanced our understanding of pathogen transmission dynamics in populations with heterogeneous contact (1–3). Network analysis gained traction with the field of veterinary epidemiology over a decade ago and has often been applied to livestock and wildlife populations in an attempt to unravel the impact of contact heterogeneity on the spread of pathogens (4–9). These advancements have led to greater knowledge surrounding potential risks for disease spread, which ultimately support decision-making pertaining to resource allocation for surveillance, management, and control strategies (10–12).

Although social network approaches provide a robust framework to study a variety of systems, they can fall short of capturing complexity associated with interactions that are commonly considered in veterinary epidemiology. In many contexts considering the role of different types

of contact (e.g., different types of social interactions, different types of movement between farms or interactions between different species) can have a significant impact on our understanding of how infectious diseases spread (13–15). Multilayer networks facilitate such an approach by including multiple network layers to more explicitly represent features of natural systems (16, 17). In traditional contact networks, or disease-relevant social networks, nodes represent individuals or populations, and edges represent disease-relevant contacts between the nodes. In multilayer networks, nodes are organized into layers, and edges can connect nodes in the same layer (*intralayer edges*) or nodes in different layers (*interlayer edges*) (**Figure 1**).

The separation of layers within the multilayer framework allows for the coupling of dynamical processes across and within layers and has consequently revealed phenomenon unattainable through traditional network representations (18). For example, the multilayer network framework has been used to capture epidemiological processes contributing to our understanding of the influence of information spread (19, 20), social support on infectious disease transmission (19–22), the role of different species in multi-host infections (23, 24), and the role of different modes of transmission in infectious disease dynamics (25, 26).

The purpose of this review is to highlight the potential uses of multilayer networks in veterinary epidemiology. The review is divided into four main sections. The first describes key terms and techniques commonly used in multilayer network analysis. We then review the use of multilayer models in human and veterinary epidemiology. We provide an example using U.S. swine networks representing contact through swine shipments and spatial proximity. Finally, we discuss important considerations when using the approach in an epidemiological context and outline some key research questions that multilayer network approaches will help veterinary epidemiologists address.

MULTILAYER NETWORK METHODOLOGY

Terminology

The power of multilayer networks lies in their flexibility to characterize multiple types of interactions not possible using a traditional monolayer network approach. In monolayer networks, *edges* (or links) represent connections between *nodes* that can be directed or undirected. For example, networks may describe social associations (undirected edges) among wild animals (each individual being a node) or movements (directed edge) from one farm to another (each farm being a node). Multilayer networks also consist of *nodes and edges*, but the nodes exist in separate *layers*, representing different forms of interactions, which connect to form an *aspect* (16, 17). Aspects, or stacks of layers, can be used to represent different types of contacts, spatial locations, subsystems, or points in time. The edges between nodes in the same layer of an aspect are called *intralayer connections,* whereas edges between nodes in different layers are *interlayer connections* (17, 18, 23).

There are two main types of multilayer networks, *multiplex networks* and *interconnected networks* (17, 27). In *multiplex networks,* interlayer edges can only connect nodes that represent the same actor in different layers. Therefore, multiplex networks typically represent sets of interactions between the same (or a similar set) of entities (e.g., individuals, farms). In *interconnected networks*, interlayer edges can connect between different actors, and therefore different layers typically represent different entities (e.g., individuals of different species, or farms in different production systems) (**Figure 1**). Thus, the structure of interlayer edges can be used to distinguish different types of multilayer network. When interlayer edges can only link nodes to nodes representing the same entity (the same individual animal or farm) in different layers, the network is classified as a *multiplex network* (28, 29). When interlayer edges can link nodes representing one entity to nodes connecting others in different layers then the network is classified as an *interconnected network*.

*Multi-relational network*s are an example of a multiplex network (30). In multi-relational networks, layers may represent the same population of individuals but with different forms of contact, which is advantageous for representing different modes of transmission. For example, one layer may represent direct contact in which edges represent the shipment of animals between farms, and the other may represent indirect contact through edges representing a shared source of feed. Another example of a multiplex network is a *temporal network* in which each node is connected to itself over discrete layers that represent time periods, but the connections between individuals within a layer represent interactions captured during that duration of time (16, 17, 30). Understanding variation in the temporality of disease processes can be critical to the application of intervention activities as well as providing useful information surrounding potential sources of infection.

Extending Centrality Measures to Multilayer Networks

Centrality is often used as a measure of an individual's importance in a network and as proxy for its role in the transmission of infection (31). Measures of centrality include local measures such as *degree* and *strength* that take into account only immediate neighbors in the network (31), global measures such as *closeness* and *betweenness centrality* that take into account the entire network structure (31), and intermediate measures such as *eigenvector* (32)*, Katz and PageRank centralities* (33) that account for some indirect connections when calculating the influence of an individual. In monolayer networks, centrality measures have been used to identify individuals with disproportionately large numbers of contacts that serve as potential super-spreaders (34) or can be crucial cut-points (35) or capacitors (36) in the spread of infection.

The multilayer network approach allows for flexibility to capture an individual's engagement in contact across a variety of disease-relevant contexts by extending the suite of centrality measures to consider interactions within and across layers. *Multidegree* is a vector of the connectedness of an individual in each layer of a multiplex network, and the same vector of centralities can be used for other measures (37). Quantifying the centrality of nodes for multiple

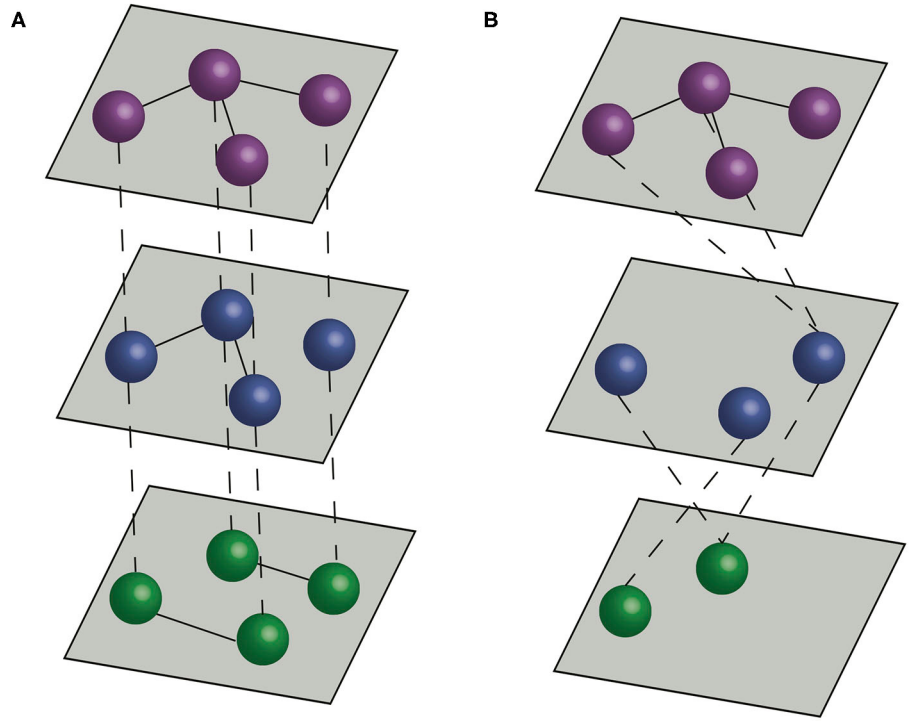


FIGURE 1 | Multilayer networks. Dashed lines represent interlayer connections, and solid lines represent intralayer connections. **(A)** A multiplex network formed by three layers, with interlayer edges connecting the same individual across layers. **(B)** An interconnected network formed by three layers, with interlayer edges connecting different individuals across layers.

layers makes it possible to consider how its connectedness is distributed across layers and can provide nuance in identifying which individuals might be most important to the spread of infection through different transmission modes. *Versatility* provides a single measure of a node's importance across multiple layers and considers the full multilayer structure (38). Various versatility metrics can be implemented for *betweenness*, *eigenvector,* and *PageRank* centralities (39), and can be calculated using the *MuxViz* software (40). For centrality metrics that are based on paths within a network, such as betweenness, it readily apparent how a multi-layer index that allows a path to traverse the network via several different layers could better capture a node's importance when there are multiple transmission modes. Individuals or farms that are not especially well connected in any one layer may have the highest versatility if they are well connected across multiple layers.

Mulitplex Neighborhoods and Relevance

It is also possible to calculate the importance of particular layers within multiplex networks. The *neighborhood* of an individual in a single or specified set of layers is the number of actors connected to an actor (or node) in that layer (or set of layers). From this it possible to calculate the *exclusive neighborhood*, the number of nodes directly connected to a focal node only in that layer or set of layers, and the *connective redundancy* of a layer (or set of layers) which is $1 - \frac{neighborhood}{total\ degree}$. Finally, the *relevance* of a layer is the percentage of neighbors present in a specified set of layers, and the *exclusive relevance* is the percentage of neighbors only present in that set of layers. These measures can be calculated using the *multinet* package (41) in R (42). They can be used to provide some indication of the role of different layers in a multiplex network, and in epidemiological context would be most useful in identifying layers that are especially important to transmission, especially in spreading infection to parts of the population that are less well connected in general.

Extending Community Detection Methods to Multilayer Networks

Often, nodes within a network are clustered. Nodes that are directly connected are more likely to share mutual connections (*transitivity*) and networks can often be subdivided into *communities* (or modules) in which within-community connections are much more frequent than connections between individuals in different communities. The strength of these subdivisions is measured using *modularity* and can have important implications for disease transmission. For example, networks with higher levels of *modularity* tend to have a slower spread of infectious disease (43). Communities in multilayer networks are defined in a similar manner, but

can account for variation in connectivity across layers (44). Frequently used examples of multilayer community-detection algorithms include multislice modularity maximization (44, 45), which maximizes the modularity quality over the network partitions by comparing the total edge weights in an observed network to the total expected edge weights in a "null network" (45), and Infomap, which maximizes the map equation by identifying cluster structures in a network and minimizing the description length of a random walker on a network (39, 46). Community detection in multilayer networks might be useful in taking into account multiple transmission routes (i.e., different types of contact) while identifying epidemiologically relevant clusters of individuals that could represent single units for management interventions. It could also be used to identify clusters of individuals that play a key role in disease spread through multiple routes of transmission, but at different time points.

Compartmental Models on Multilayer Networks

Mathematical modeling has long been an important tool in veterinary epidemiology, principally in the form of compartmental models (47). These approaches model the transition of individuals between disease states with examples including the widely used SI (susceptible-infected), SIR (susceptible-infective-recovered), SIS (susceptible-infective-susceptible), and SEIR (susceptible-exposed-infectious-recovered) models. In general, compartmental models on networks tend to be individual-based (5, 48, 49), but veterinary epidemiological studies have developed population-based network models (50), which are often more suitable for studying livestock populations. Compartmental models have already been applied to study the spread of infectious disease in multilayer networks (21, 22, 51, 52), and can frequently provide additional insights into infectious disease dynamics. Methods for modeling infectious disease transmission on multilayer networks are similar to those developed for compartmental metapopulation models of disease spread but generally support higher levels of complexity than metapopulation models, as they allow for the integration of multiple modes of contact within and between population and other interconnecting processes (53). It has been shown that, when interlayer edges connect individuals in different, discrete populations and intralayer edges connect individuals within each population, certain distributions of interlayer vs. intralayer edges can cause outbreaks in the system as a whole which would not occur in any single population (layer) within the system. Further, under certain conditions, the epidemic threshold of the whole system may be smaller than the epidemic threshold of its parts (54). These additional insights can be important in exploring the effects of interventions strategies aimed at different subpopulations or the effects of multiple spreading processes, such as disease awareness or vaccination behavior (22, 55), which can continue to advance our understanding of the influence of complex contact structures on infectious disease dynamics.

PREVIOUS USES OF MULTILAYER NETWORKS IN EPIDEMIOLOGY

The scientific study of multilayer networks is a burgeoning area of research, particularly the development of theoretical epidemiological models and its application to human epidemiology. Although its use in veterinary contexts is still limited, here we outline key areas of research that have been pursued in theoretical and empirical studies (in both humans and animals) and highlight how multilayer networks might be applied to veterinary epidemiology.

Different Routes of Infection

Multilayer networks can be usefully applied in contexts where a pathogen can be transmitted through multiple modes or pathways of infection (12), as the multiplex approach provides a framework to account for multiple transmission probabilities. Considering the presence of multiple transmission modes can influence the efficacy of targeted interventions, particularly if nodes were traditionally targeted according to their degree in only one layer (25, 26, 56). This has implications for situations where data, networks, and resultant optimal control strategies are only available for one mode of transmission, leading to overconfidence in the efficacy of control.

In the context of veterinary epidemiology, animal movements are typically considered the most effective transmission mode between farms (direct contacts) (57). However, other infection mechanisms might play an important role such as wind-borne spread and fomites disseminated through contaminated clothes, equipment, and vehicles by personnel (indirect contacts) (58–60). Ignoring one mode of transmission could lead to inaccurate farm risk predictions and ineffective targeted surveillance. This has been demonstrated in a network analysis that considered both direct (cattle movements) and indirect (veterinarian movements) contacts to reveal that indirect contact, despite being less efficient in transmission, can play a major role in spread of a pathogen within a network (13).

In another example, Stella et al. (51) used an "ecomultiplex model" to study the spread of *Trypanosoma cruzi* (cause of Chagas disease in humans) across different mammal species. This pathogen can be transmitted either through invertebrate vectors (Triatominae or kissing bugs) or through predation when a susceptible predator feeds on infected prey or vectors. Thus, their model included two ecological/transmission layers: the food-web and vector layers. Their results showed that studying the multiplex network structure offered insights on which host species facilitate parasite spread, and thus which would be more effective to immunize in order to control the spread. At the same time, they showed how, in this system, when parasites spread occurs primarily through the trophic layer, immunizing predators hampers parasite transmission more than immunizing prey.

Furthermore, multilayer network analysis can help differentiate between different types of social interactions that may lead to disease transmission. For example, sex-related dynamics of contact networks can have important implications for disease spread in animal populations, as

seen in the spread of *Mycobacterium bovis* in European badgers (*Meles meles*) (61). The authors constructed an interconnected network that distinguished male-male, female-female, and between-sex contacts recorded during proximity loggers. Inter-layer between-sex edges and edges in the male-male layer were more important in connecting groups into wider social communities, and contacts between different social communities were also more likely in these layers.

Dynamics of Coupled Processes—the Spread of Two Pathogens

Another application of multilayer networks in epidemiology is to model the concurrent propagation of two entities through a network, such as two different pathogens co-occurring in the same population or the spread of disease awareness alongside the spread of infection. In both scenarios, the spread of one entity within the network interacts with the spread of the other, creating a coupled dynamical system. A multiplex approach can allow for each coupled process to spread through a network that is based on the appropriate type of contact for propagation (i.e., contact networks involved in pathogen transmission vs. interaction or association networks that allow information to spread). In the case of two infectious diseases concurrently spreading through a network, a multiplex approach can be particularly useful if infection of a node by pathogen *A* alters the susceptibility to pathogen *B*, or if coinfection of a node influences its ability to transmit either pathogen. For example, when infection by one pathogen increases the likelihood of becoming infected by another pathogen, it could theoretically facilitate the spread of a second pathogen and thus alter epidemic dynamics (62). This type of dynamic is likely to widespread in wild and domestic animals due to the importance of co-infection in affecting infectious disease dynamics by influencing the replication of pathogens within hosts (63). However, when there is competition or cross-immunity, the spread of one pathogen could reduce the spread of a second pathogen (64). For example, this type of dynamic could be expected for pathogens strains characterized by partial cross-immunity, such as avian influenza (65), or microparasite-macroparasite coinfections in which infection with one parasite reduces transmission of a second, such as infection with gastrointestinal helminths reducing the transmission of bovine tuberculosis in African buffalo (*Syncerus caffer*) (66). Similar "within-node" dynamics could be important at a farm-level in livestock movement networks. For example, the detection of a given pathogen infection in a farm might cause it to be quarantined, thus reduce its susceptibility and ability to transmit other pathogen infections.

Dynamics of Coupled Processes—Interactions Between Transmission Networks and Information/Social Networks

For coupled processes involving a disease alongside a social process (i.e., spread of information or disease awareness), we might expect that the spread of the pathogen will be associated with the spread of disease awareness or preventative behaviors such as mask-wearing, and in these cases theoretical models suggest that considering the spread of disease awareness can result in reduced disease spread (67). A model was presented by Granell et al. (19), which represented two competing processes on the same network: infection spread (modeled using a Susceptible-Infected-Susceptible compartmental model) coupled with information spread through a social network (an Unaware-Aware-Unaware compartmental model). The authors used their model to show that the timing of self-awareness of infection had little effect on the epidemic dynamics. However, the degree of immunization (a parameter which regulates the probability of becoming infected when aware) and mass media information spread on the social layer did critically impact disease spread (19). A similar framework has been used to study the effect of the diffusion of vaccine opinion (pro or anti) across a social network with concurrent infectious disease spread. The study showed a clear regime shift from a vaccinated population and controlled outbreak to vaccine refusal and epidemic spread depending on the strength of opinion on the perceived risks of the vaccine. The shift in outcomes from a controlled to uncontrolled outbreak was accompanied by an increase in the spatial correlation of cases (20). While models in the veterinary literature have accounted for altered behavior of nodes (imposition of control measures) as a result of detection or awareness of disease (68), it is not common for awareness to be considered as a dynamic process that is influenced by how each node has interacted with the pathogen (i.e., contact with an infected neighbor). For example, the rate of adoption of biosecurity practices at a farm, such as enhanced surveillance, use of vaccination, or installation of air filtration systems, may be dependent on the presence of disease in neighboring farms or the farmers' awareness of a pathogen through a professional network of colleagues.

There is also some evidence that nodes that are more connected in their "social support" networks (e.g., connections with family and close friends in humans) can alter network processes that result in negative outcomes, such as pathogen exposure or engagement in high-risk behaviors (22). In a case based on users of injectable drugs, social connections with non-injectors can reduce drug-users connectivity in a network based on risky behavior with other drug injectors (69). In a model presented by Chen et al. (22), a social-support layer of a multiplex network drove the allocation of resources for infection recovery, meaning that infected individuals recovered faster if they possessed more neighbors in the social support layer. In animal (both wild and domesticated) populations, this concept could be adapted to represent an individual's likelihood of recovery from, or tolerance to, infection being influenced by the buffering effect of affiliative social relationships (70). For domestic animals, investment in certain resources at a farm level could influence a premise's ability to recover (e.g., treatment) or onwards transmission of a pathogen (e.g., treatment or biosecurity practices). Sharing of these resources between farms could be modeled through a "social-support" layer in a multiplex, for example, where a farm's transmissibility is impacted by access to shared truck-washing facilities.

Multi-Host Infections

Multilayer networks can be used to study the features of mixed species contact networks or model the spread of a pathogens in a host community, providing important insights into multi-host pathogens (12). Scenarios like this are commonplace at the livestock-wildlife interface and therefore the insights provided could be of real interest to veterinary epidemiology. In the case of multi-host pathogens, intralayer and interlayer edges represent the contacts between individuals of the same species and between individuals of different species, respectively. They can therefore be used to identify bottlenecks of transmission and provide a clearer idea of how spillover occurs. For example, Silk et al. (24) used an interconnected network with three layers to study potential routes of transmission in a multi-host system. One layer consisted of a wild European badger (*Meles meles)* contact network, the second a domesticated cattle contact network, and the third a layer containing badger latrine sites (potentially important sites of indirect environmental transmission). No intralayer edges were possible in the latrine layer. The authors demonstrated the importance of these environmental sites in shortening paths through the multilayer network (for both between- and within-species transmission routes) and showed that some latrine sites were more important than others in connecting together the different layers. Pilosof et al. (23) presented a theoretical model, labeling the species as focal (i.e., of interest) and non-focal, showing that the outbreak probability and outbreak size depend on which species originates the outbreak and on asymmetries in between-species transmission probabilities.

Similar applications of multilayer networks (see **Supplementary Material**) could easily be extended to systems where two or more species are domesticated animals, as well. Examples of these could be the study of a pathogen such as Bluetongue virus, which affects both cattle and sheep (71), or foot-and-mouth disease virus, which infects cattle, sheep, and pigs (60). In such cases, each species can be represented by a different level in the network, and interlayer edges are made possible as a result of mixed farms (i.e., cattle and sheep), different species from different farms grazing on the same pasture, or for other types of indirect contacts such as the sharing equipment or personnel.

Overall, multilayer approaches provide an elegant way to analyze cross-species transmission and spillover, including for zoonotic pathogens across the human-livestock-wildlife interface. They can be used to simultaneously model within-species transmission, identify heterogeneities among nodes in their tendency to engage in between-species contacts relevant for spillover and spillback, and better predict the dynamics of spread prior and subsequent to cross-species transmission events, which may contribute to forecasting outbreaks in target species. Measures of multilayer network centrality in this instance could be used to extend the superspreader concept into a community context; individuals that are influential in within-species contact networks and possess between-species connections might be predicted to have a more substantial influence on infectious disease dynamics in the wider community.

CASE STUDY: MULTIPLEX NETWORKS IN THE U.S. COMMERCIAL SWINE INDUSTRY

To demonstrate the utility and application of multilayer network analysis, we provide an example from the commercial swine industry in the United States. Our objective is to cement the concepts presented in this review with a real-world example, and to demonstrate how a multi-layer approach can enhance insights on the identity of high-risk nodes (highly connected farms that have greater exposure or are potential super-spreaders) and the architecture and modularity of networks when multiple modes of contact are considered. In this example, we calculate centrality metrics and identify communities using data from 1,544 farms belonging to two swine companies from production systems that have been previously described (72). Both companies are vertically integrated, meaning that different phases of production occur at different farms (gestation and farrowing at *sow farms*, rearing of weaned piglets at *nursery farms*, and fattening pigs for the market at *finishing farms*). We created a multiplex network with two layers to account for multiple modes of transmission-relevant contact. Intralayer edges in one layer consisted of animal movement between farms as it is a known pathway of pathogen transmission between farms. The second layer consisted of predicted contacts arising from spatial proximity (threshold at < 5 km), because it has also been postulated that local area spread occurs via windborne spread or indirect contacts, such as shared personnel, trucks or equipment, for several important swine diseases (58, 73). For example, spatial proximity networks based on a 5 km threshold have been shown to be associated with the occurrence of porcine reproductive and respiratory syndrome virus (72). There are additional nuances to both spatial proximity (e.g., wind direction, climatic factors, vegetation (74), and animal movements (temporality, directed vs. undirected, etc.) that should be accounted for in a rigorous analysis of transmission within swine systems, but we have simplified these to create a clearer conceptual illustration of multi-layer networks. From an initial visual assessment of **Figure 2**, it is apparent that either layer alone would misrepresent connectivity patterns. Spatial proximity overestimates the fragmentation of the network across space, while animal movements underrepresent local connections.

We quantified the centrality of each node in the multiplex network, in each single-layer network, and the overall aggregated network using MuxViz v2.0.1 (40). In this analysis, we focused on degree, strength, and eigenvector centrality. Because the spatial proximity network was denser than the movement network, we re-scaled the edge weights such that the sum weight of all edges was equal to one in both networks. This helped ensure that the spatial proximity layer, which had higher density, was not excessively dominant over the movement layer which contained many fewer edges. In practice, the relative weighting of edges in different layers should be subject to a sensitivity analysis or tested with data (see *Points of considerations* section below), as this choice can influence multilayer metrics and communities. However, we used a simple re-scaling approach here to demonstrate multilayer concepts.

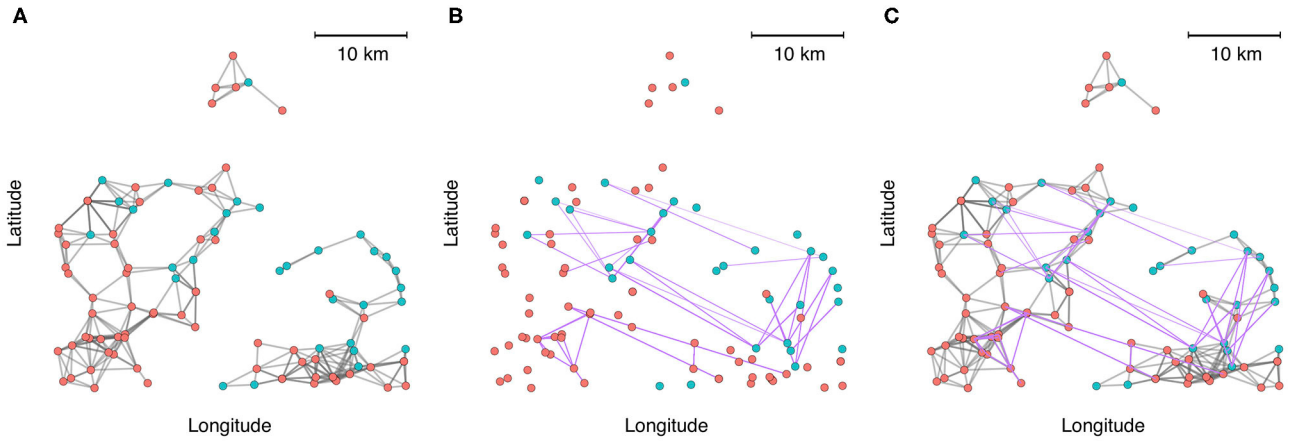


FIGURE 2 | A zoomed-in subset of 100 farms plotted into geographic space, where node color represents each production system. **(A)** Connections between farms based on spatial proximity (<5 km). **(B)** Connections between farms based on animal movements. **(C)** Aggregate visualization of the spatial proximity and movement network, where edges are colored according to the type of contact (gray = spatial proximity, purple = animal movement).

Outputs of this analysis were visualized as an annular plot in which a node appears in the same position in each ring of the plot (**Figure 3**). **Figure 3A** shows the annular visualization of node centrality for the subset of farms shown in **Figure 2**, with each segment representing a different node in the multiplex network and each ring representing the network layers. Across all three metrics shown here, it is clear that there is a variable correlation in centrality across the single-layer and multilayer networks (Spearman correlation coefficients range from −0.18 to 1.0, **Figure 3**). In particular, we see that farms with high strength or eigenvector centrality in the movement network are not necessarily the same farms that have high values of these measures in the spatial proximity network. For targeted disease control, the selection of key nodes based on a single layer could therefore be misleading.

Targeted disease control in livestock industries, especially as an outbreak response strategy, can also rely on defining control zones around infected premises, with strict control measures applied to farms within these zones. A related strategy, zonation, relies on defining regions of a country as disease-free for the purposes of international trade. An alternative approach to defining zonation and control zones called compartmentalization has also been proposed. A compartment is defined as a subpopulation of interlinked premises (such as a swine production system) with a common health status with respect to a specific disease, limited contact with premises outside the compartment, and for which surveillance, control, and biosecurity measures have been established for the purposes of trade (75). For a pathogen with multiple modes of transmission, it would be logical to define compartments based on connectivity in a multiplex network.

Here, we demonstrate the use of the Infomap multilayer community finding algorithm to define such compartments. Communities are thus defined as groups of farms that are in greater contact with one another than with farms outside of their communities. Critically, here contact between farms of the same community can either be through animal movement or spatial proximity. In the Infomap analysis, each node is assigned to a community in both the movement and spatial proximity layer, and some communities span both layers (**Figure 4A**). Our Infomap analysis identified numerous communities. If we map out the distribution of five largest communities in geographic space (**Figure 4B**), we see that each community generally includes several groups of farms that cluster tightly together in space, reflecting connectivity in the spatial proximity layer. However, different spatial clusters can occur within the same community if they are interlinked in the movement layer. This approach could thus be used to define groups of epidemiologically linked farms as compartments for pathogens with multi-modal transmission. From a disease control perspective, these compartments could be used to define high-risk (pathogen detected within the compartment) or low-risk farms (pathogen not yet detected in the compartment). Additional hypotheses could also be tested about transmission, such as the extent to which community membership influences pathogen diversity (i.e., do different communities have genetically distinct variants of a pathogen?).

IMPORTANT CONSIDERATIONS WHEN USING MULTILAYER NETWORKS

Data Collection

Appropriate collection of network data is a fundamental challenge in the design of network studies, and it is important to ensure enough data is collected to provide a realistic insight into the study system (76–79). This problem is enhanced when using network modeling approaches in epidemiology, where missing edges can result in substantial underestimates of outbreak sizes (80). On the other hand, a similar but opposite problem might arise when the lack of a transmission-relevant contact data triggers the use of imprecise proxies such as shared contractors (i.e., two farms are considered connected when they use the

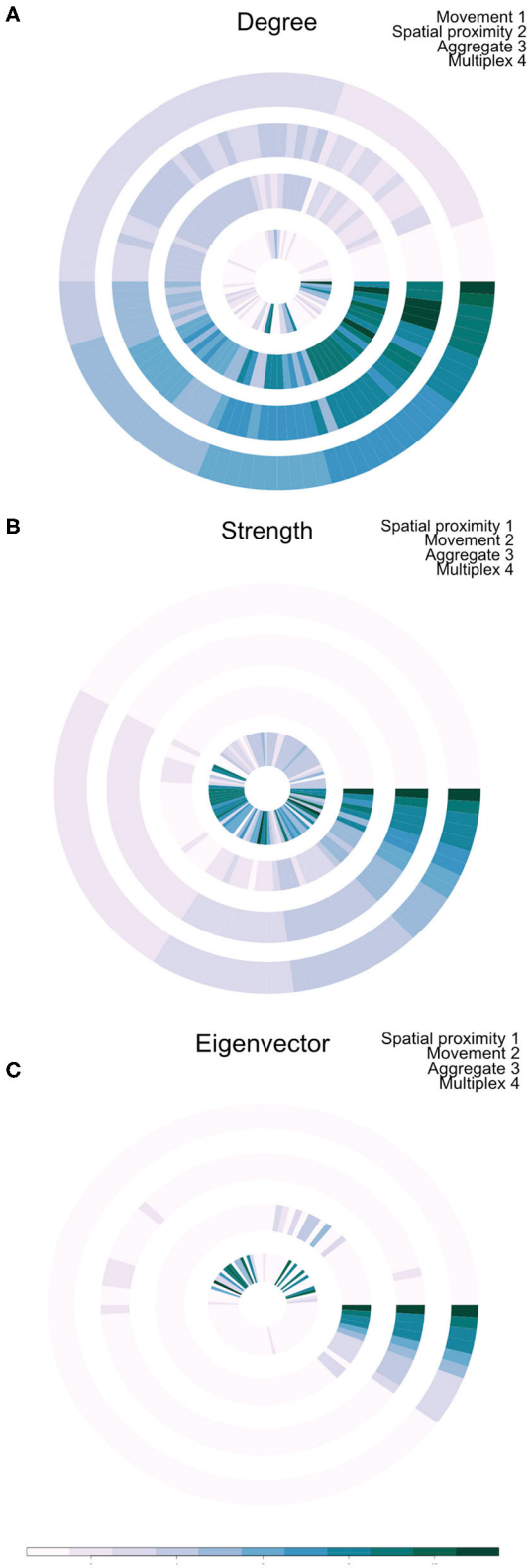


FIGURE 3 | Annual visualization of node centrality metrics: **(A)** degree, **(B)** strength, and **(C)** Eigenvector centrality. Each node appears in the same position in each ring. Darker colors indicate nodes with higher centrality values. Each ring visualizes node centrality when measured in each layer separately (inner two rings), in the aggregate network (ring 3), and using the multiplex version of the metric (outer ring 4).

same company for feed delivery, milk trucks, etc.): this could lead to overestimating the potential epidemic spread (13). For studies of infectious disease, it is typically necessary to construct networks over a time period relevant to the transmission of the infection of interest to increase the accuracy of network-based inference (81–83), which may provide time constraints on when data can be collected. Using multilayer network approaches can exacerbate this difficulty if it requires researchers to collect data on more different types of contact or interaction for multiplex networks, or potentially across more species or fomites for interconnected networks. Therefore, the feasibility of collecting sufficient data is an important consideration when weighing up whether to use multilayer approaches. Multilayer networks might be most naturally applied in wildlife epidemiology studies in which multiple different types of interactions between individuals are already recorded (84–86). In livestock context, while animal movements data have been regularly collected and analyzed in several countries in the past two decades (11, 57, 59, 87–89), challenges in adopting multilayer approach might arise for (i) countries where the collection of movement data and other industry-related information (e.g., farm location) is not mandatory, in particular developing countries (76), and (ii) including non-animal movements related potential infectious contacts, which data are often scarce and temporally limited (13, 90).

Network Construction

There are also important considerations to be made when constructing multilayer networks for use in epidemiological studies. When layers consist of very different types of contacts or interactions involving distinct behaviors that are performed at different rates, it is possible that layers may differ drastically in their edge weights, and this can lead to problems with their analysis (16, 30). The same problem can also occur if sampling effort differs between layers. In multiplex networks with one layer that is much more well-connected than others, inferred transmission dynamics tend to be almost entirely controlled by the network structure of that single layer (91, 92). Therefore, in these contexts, it may be important to consider what added benefits using a multilayer approach can bring. If using a multilayer approach is still favored, then a variety of approaches are available to change the contribution of different layers (30) such as thresholding to produce unweighted or binary networks, or scaling/normalizing edge weights between layers (see our case study in this paper). In many situations, edge weights carry important information, especially when networks have a high density of connections, and incorporating edge weights can be important in network modeling of infection (93). As a result, any decision to threshold edge weights should be done with caution and be appropriate for the question being asked.

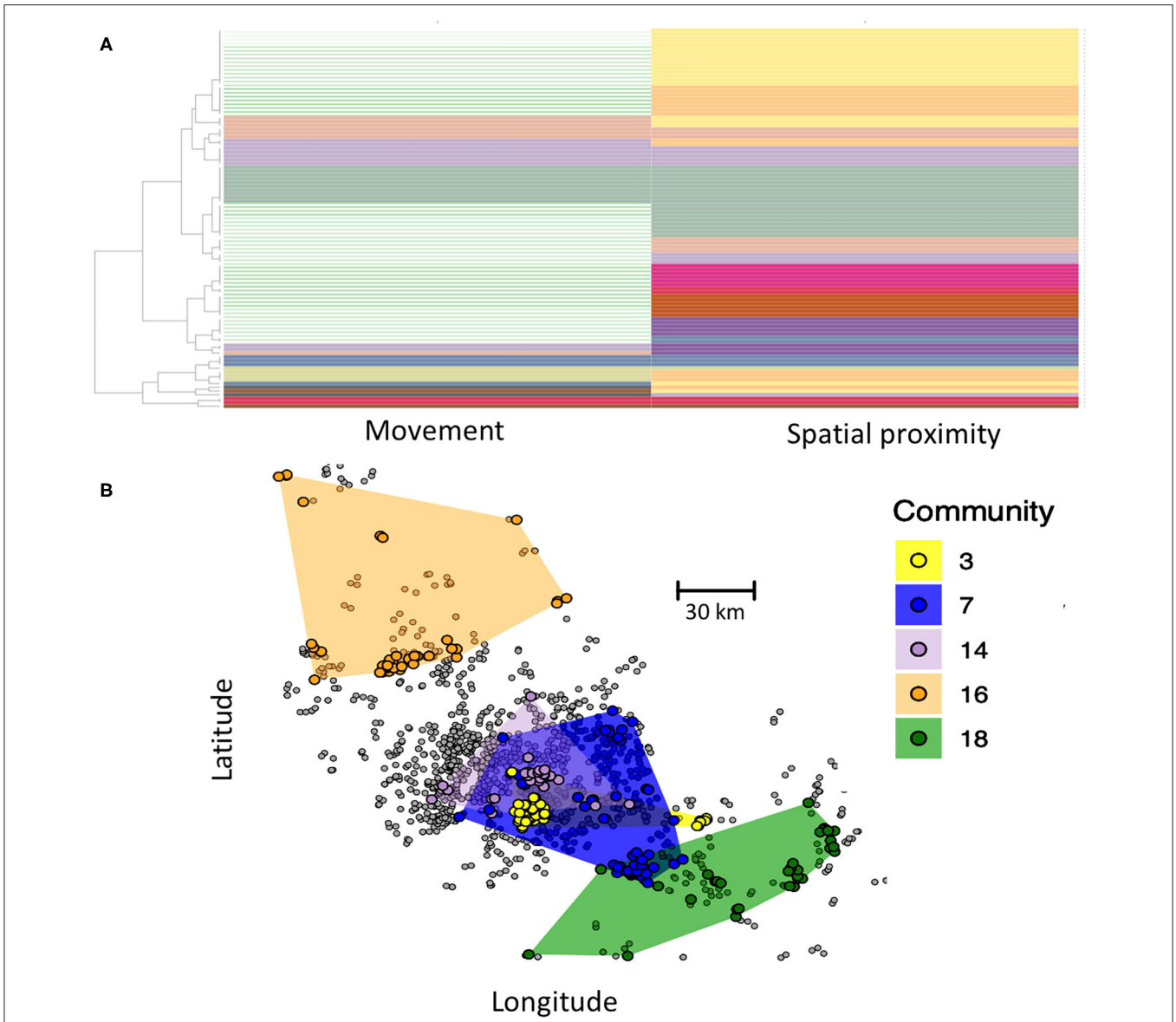


FIGURE 4 | (A) Community structure of nodes based on Infomap community detection. Each row represents a farm, and each column represents its community assignment in the movement and spatial proximity networks, with some communities spanning both layers (same color in both columns). Rows with no color in the movement network are farms that only occurred in the spatial proximity network (no animal movements recorded) and thus did not have a community assignment in the movement layer. **(B)** Map of the five largest communities, with color of circles representing the community membership of farms. Spatial distribution of each community is shown by the colored polygons, which were determined by creating a minimum convex polygon around each community's farms. Farms in gray were not part of any of the largest five communities. Inter-layer edge weights were set to 0.01.

A related consideration in multiplex networks is whether there is any redundancy between different layers (e.g., sets of intra-layer connections that are closely correlated with each other and represent the same set of ties). There are now a number of approaches available to calculate redundancy between layers in multiplex networks (94) that can provide valuable insights into the importance of taking a multilayer approach.

An important independent consideration in multiplex network studies is how interlayer edges should be weighted (16). While in interconnected networks, both intra- and interlayer network studies have natural weights, interlayer edge weights in multiplex networks are less intuitive as they typically connect the same actor to itself in different layers (whether this is an individual in a contact network or a farm in a livestock movement network). Epidemiological research offers an opportunity to provide interlayer edge weights with meaningful values in multiplex networks, especially when network modeling approaches are used. In these cases, interlayer edges can be

used to represent the probability of being infected through one layer, causing an individual to be infectious in the second layer. For example, in a multiplex social network that included layers representing biting interactions (with knowledge that these could provide a transmission route) and close contact (as a proxy for aerosol transmission), interlayer edges could be weighted by the probability that an individual infected through being bitten by an infectious neighbor could subsequently transmit the infection through the contact network. A similar approach could be used for other multiplex networks (such as the between-farm networks in our case study). When these probabilities are unknown, then a sensitivity analysis on interlayer edge weights could be used to test the robustness of any conclusions drawn to these values, or alternatively, when empirical disease data is available, it might be possible to estimate these probabilities using an appropriately implemented network model.

Points of Consideration

There remain some practical limitations in the analysis of multilayer networks (30). While methods for calculating descriptive metrics for multiplex networks have been widely developed and can be implemented using software packages in R (42) and Python (95), methods for the analysis of other multilayer networks (e.g., interconnected networks) are much less accessible. Therefore, when analyzing or modeling interconnected networks, it may be important to review the options available or feel confident in applying the calculations or algorithms required in the absence of ready-built functions. A similar consideration needs to be made when applying randomization-based analyses in multilayer networks. Especially for wildlife-based studies, comparison of multilayer networks to suitable permutations is likely to be important (30), and when conducting randomizations for multilayer networks it is important to consider both the research question being asked and any additional network features that arise as an outcome of the multilayer network structure of the data (17, 30).

OPPORTUNITIES FOR FUTURE USE OF MULTILAYER NETWORKS IN VETERINARY EPIDEMIOLOGY

An exciting area for future application of multilayer networks in veterinary epidemiology includes the delineation of functional relationships between livestock systems for the implementation of subpopulation management strategies for World Organization for Animal Health (OIE)-listed diseases, as mentioned in the U.S. commercial swine industry example in a previous section. Establishing disease-free status throughout a country can be a difficult and timely undertaking. As such, the concepts of "zoning" and "compartmentalization" were developed by the OIE to recognize animals with different health statuses based on the geographical location (zoning) or based on management practices (compartmentalization) to facilitate the continuation of trade. A multilayer approach can be used to identify such subpopulations or estimate the risk of disease spread between subpopulations given different modes of transmission, which can guide the designation and maintenance of a subpopulation's disease-free status.

Moreover, multilayer networks can advance our knowledge of how temporal dynamics of network structure influence infectious disease spread. Contact networks of both wild and domestic animals are inherently dynamic (82, 96), and information contained in these contacts can change the rate of pathogen spread, as well as the efficacy of control strategies based on static networks. Many monolayer contact networks that incorporate temporality assume some level of aggregation of contacts over a period of time, such as calculating the mean or total edge weight for all edges over all time-steps, then use traditional techniques to characterize network properties (45). Others use traditional monolayer approaches to characterize network properties within each time-step then analyze the changes over the time-ordered layers (45). However, either approach results in loss of information and the ability to understand more complex interactions such as simultaneous interactions of multiple modes of contact and their evolution over time (97). It is prudent to note that the temporal analysis of an epidemiological process should consider the natural history of the pathogen under investigation in order to reflect the underlying epidemiological processes appropriately. Despite increases in research activity addressing the influence of temporality on complex systems, there is much to uncover regarding the impact of duration, concurrency, order of network properties, especially pertaining to disease transmission.

Multilayer network approaches are likely to be especially valuable at the (human-) livestock-wildlife interface (24), where identifying multi-host dynamics of pathogens is particularly important (98). Properly implemented multilayer network models will make it possible to better quantify the role of wildlife reservoirs of infection and estimate the rate of spillover from wildlife to livestock and vice versa. Taking these approaches may facilitate the identification of bottlenecks to transmission that can represent targets for management interventions or promote an understanding of the characteristic of individual animals or premises that play disproportionate roles in the spread or maintenance of infection in a community context. This extends naturally to encompass vector-borne transmission, especially for multi-host vector-borne diseases such as yellow fever, Lyme disease, and West Nile virus. A multilayer network approach could be used to unravel vital questions surrounding the management of such pathogens by incorporating factors that influence transmission, such as vector preference and host transmissibility.

CONCLUSIONS

In this paper, we provide an overview of the early use multilayer networks in human and veterinary epidemiology. From the dynamics of coupled processes, such as information spread and disease transmission, to multi-host transmission, multilayer networks have been used to analyze a range of complex epidemiological systems that have been challenging to study in monolayer networks. Despite the caveats associated with their use, multilayer networks show promise in providing a powerful

framework for furthering our understanding of the complex interactions that influence disease transmission dynamics in veterinary medicine.

AUTHOR CONTRIBUTIONS

AK, GR, MS, and KV contributed conception and design of the study and wrote sections of the manuscript. KV performed the analysis. All authors contributed to manuscript revision and approved the submitted version.

FUNDING

This study was partially funded by the Swine Health Information Center (SHIC). Funding was also provided by the joint NIFA-NSF-NIH Ecology and Evolution of Infectious Disease award 2019-67015-29918, the Agriculture and Food Research Initiative Competitive grant no. 2018-68008-27890 from the USDA National Institute of Food and Agriculture, the University of Minnesota, and the University of Exeter. GR was supported by BBSRC GRANT BB/P010598/1.

ACKNOWLEDGMENTS

We would like to express our thanks to the swine producers and practitioners for sharing their data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.00596/full#supplementary-material

REFERENCES

1. Keeling MJ, Eames KT. Networks and epidemic models. *J R Soc Interface.* (2005) 2:295–307. doi: 10.1098/rsif.2005.0051
2. Pastor-Satorras R, Vespignani A. Epidemic dynamics and endemic states in complex networks. *Phys Rev E.* (2001) 63:066117. doi: 10.1103/PhysRevE.63.066117
3. Newman M. *Networks*. 2nd ed. Oxford, UK: Oxford University Press (2018).
4. Martínez-López B, Perez AM, Sánchez-Vizcaíno JM. Social network analysis. Review of general concepts and use in preventive veterinary medicine. *Transbound Emerg Dis.* (2009) 56:109–20. doi: 10.1111/j.1865-1682.2009.01073.x
5. Dubé C, Ribble C, Kelton D, McNab B. Introduction to network analysis and its implications for animal disease modelling. *OIE Rev Sci Tech.* (2011) 30:425–36. doi: 10.20506/rst.30.2.2043
6. Craft ME. Infectious disease transmission and contact networks in wildlife and livestock. *Philos Trans R Soc Lond Ser B Biol Sci.* (2015) 370. doi: 10.1098/rstb.2014.0107
7. VanderWaal KL, Atwill ER, Hooper S, Buckle K, McCowan B. Network structure and prevalence of Cryptosporidium in Belding's ground squirrels. *Behav Ecol Sociobiol.* (2013) 67:1951–9. doi: 10.1007/s00265-013-1602-x
8. Rushmore J, Caillaud D, Hall RJ, Stumpf RM, Meyers LA, Altizer S. Network-based vaccination improves prospects for disease control in wild chimpanzees. *J R Soc Interface.* (2014) 11:20140349. doi: 10.1098/rsif.2014.0349
9. VanderWaal KL, Atwill ER, Isbell LA, McCowan B. Quantifying microbe transmission networks for wild and domestic ungulates in Kenya. *Biol Conserv.* (2014) 169:136–46. doi: 10.1016/j.biocon.2013.11.008
10. Kiss IZ, Green DM, Kao RR. The network of sheep movements within Great Britain: network properties and their implications for infectious disease spread. *J R Soc Interface.* (2006) 3:669–77. doi: 10.1098/rsif.2006.0129
11. Bajardi P, Barrat A, Savini L, Colizza V. Optimizing surveillance for livestock disease spreading through animal movements. *J R Soc Interface.* (2012) 9:2814–25. doi: 10.1098/rsif.2012.0289
12. Silk MJ, Hodgson DJ, Rozins C, Croft DP, Delahay RJ, Boots M, et al. Integrating social behaviour, demography and disease dynamics in network models: applications to disease management in declining wildlife populations. *Philos Trans R Soc B Biol Sci.* (2019) 374:20180211. doi: 10.1098/rstb.2018.0211
13. Rossi G, De Leo GA, Pongolini S, Natalini S, Zarenghi L, Ricchi M, et al. The potential role of direct and indirect contacts on infection spread in dairy farm networks. *PLoS Comput Biol.* (2017) 13:e1005301. doi: 10.1371/journal.pcbi.1005301
14. Godfrey SS, Bull CM, James R, Murray K. Network structure and parasite transmission in a group living lizard, the gidgee skink, *Egernia stokesii*. *Behav Ecol Sociobiol.* (2009) 63:1045–56. doi: 10.1007/s00265-009-0730-9
15. Grear DA, Luong LT, Hudson PJ. Network transmission inference: host behavior and parasite life cycle make social networks meaningful in disease ecology. *Ecol Appl.* (2013) 23:1906–14. doi: 10.1890/13-0907.1
16. Pilosof S, Porter MA, Pascual M, Kéfi S. The multilayer nature of ecological networks. *Nat Ecol Evol.* (2017) 1:0101. doi: 10.1038/s41559-017-0101
17. Kivela M, Arenas A, Barthelemy M, Gleeson JP, Moreno T, Porter MA. Multilayer networks. *J Complex Netw.* (2014)203–71. doi: 10.1093/comnet/cnu016
18. De Domenico M, Granell C, Porter MA, Arenas A. The physics of spreading processes in multilayer networks. *Nat Phys.* (2016) 12:901–6. doi: 10.1038/nphys3865
19. Granell C, Gómez S, Arenas A. Competing spreading processes on multiplex networks: awareness and epidemics. *Phys Rev E.* (2014) 90:1–7. doi: 10.1103/PhysRevE.90.012808
20. Jentsch PC, Anand M, Bauch CT. Spatial correlation as an early warning signal of regime shifts in a multiplex disease-behaviour network. *J Theor Biol.* (2018) 448:17–25. doi: 10.1016/j.jtbi.2018.03.032
21. Wang W, Liu QH, Cai SM, Tang M, Braunstein LA, Stanley HE. Suppressing disease spreading by using information diffusion on multiplex networks. *Sci Rep.* (2016) 6:29259. doi: 10.1038/srep29259
22. Chen X, Wang R, Tang M, Cai S, Stanley HE, Braunstein LA. Suppressing epidemic spreading in multiplex networks with social-support. *New J Phys.* (2018) 20. doi: 10.1088/1367-2630/aa9cda
23. Pilosof S, Greenbaum G, Krasnov BR, Zelnik YR. Asymmetric disease dynamics in multihost interconnected networks. *J Theor Biol.* (2017) 430:237–44. doi: 10.1016/j.jtbi.2017.07.020
24. Silk MJ, Drewe JA, Delahay RJ, Weber N, Steward LC, Wilson-Aggarwal J, et al. Quantifying direct and indirect contacts for the potential transmission of infection between species using a multilayer contact network. *Behaviour.* (2018) 155:731–757. doi: 10.1163/1568539x-00003493
25. Tsai Y-S, Huang C-Y, Wen T-H, Sun C-T, Yen M-Y. Integrating epidemic dynamics with daily commuting networks: building a multilayer framework to assess influenza A (H1N1) intervention policies. *Simul Trans Soc Model Simul Int.* (2011) 87:385–405. doi: 10.1177/0037549710379481
26. Mari L, Bertuzzo E, Righetto L, Casagrandi R, Gatto M, Rodriguez-Iturbe I, et al. Modelling cholera epidemics: the role of waterways, human mobility and sanitation. *J R Soc Interface.* (2012) 9:376–88. doi: 10.1098/rsif.2011.0304
27. Silk MJ, Finn KR, Porter MA, Pinter-Wollman N. Can multilayer networks advance animal behavior research? *Trends Ecol Evol.* (2018) 33:376–8. doi: 10.1016/j.tree.2018.03.008
28. Gómez S, Díaz-Guilera A, Gómez-Gardeñes J, Pérez-Vicente CJ, Moreno Y, Arenas A. Diffusion dynamics on multiplex networks. *Phys Rev Lett.* (2013) 110:1–6. doi: 10.1103/PhysRevLett.110.028701
29. Solé-Ribalta A, De Domenico M, Kouvaris NE, Díaz-Guilera A, Gómez S, Arenas A. Spectral properties of the Laplacian of multiplex

networks. *Phys Rev E Stat Nonlinear Soft Matter Phys.* (2013) 88:1–8. doi: 10.1103/PhysRevE.88.032807
30. Finn KR, Silk MJ, Porter MA, Pinter-Wollman N. The use of multilayer network analysis in animal behaviour. *Anim Behav.* (2019) 149:7–22. doi: 10.1016/j.anbehav.2018.12.016
31. Wasserman S, Faust K. *Social Network Analysis: Methods and Applications*. Cambridge; New York, NY: Cambridge University Press (1994). doi: 10.1525/ae.1997.24.1.219
32. Ruhnau B. Eigenvector-centrality - a node-centrality. *Soc Netw.* (2000) 22:357–65. doi: 10.1016/S0378-8733(00)00031-9
33. Sergey B, Page L. The anatomy of a large-scale hypertextual Web search engine. *Comput Netw ISDN Syst.* (1998) 30:107–17.
34. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature.* (2005) 438:355–9. doi: 10.1038/nature04153
35. Vanderwaal KL, Ezenwa VO. Heterogeneity in pathogen transmission: mechanisms and methodology. *Funct Ecol.* (2016) 30:1606–22. doi: 10.1111/1365-2435.12645
36. Weber N, Carter SP, Dall SRX, Delahay RJ, McDonald JL, Bearhop S, et al. Badger social networks correlate with tuberculosis infection. *Curr Biol.* (2013) 23:R915–6. doi: 10.1016/j.cub.2013.09.011
37. Menichetti G, Remondini D, Panzarasa P, Mondragón RJ, Bianconi G. Weighted multiplex networks. *PLoS ONE.* (2014) 9:e97857. doi: 10.1371/journal.pone.0097857
38. De Domenico M, Solé-Ribalta A, Omodei E, Gómez S, Arenas A. Ranking in interconnected multilayer networks reveals versatile nodes. *Nat Commun.* (2015) 6:1–6. doi: 10.1038/ncomms7868
39. De Domenico M, Lancichinetti A, Arenas A, Rosvall M. Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Phys Rev X.* (2015) 5:1–11. doi: 10.1103/PhysRevX.5.011027
40. De Domenico M, Porter MA, Arenas A. MuxViz: a tool for multilayer analysis and visualization of networks. *J Complex Netw.* (2015) 3:159–76. doi: 10.1093/comnet/cnu038
41. Magnani M, Vega D, Dubik M. *multinet: Analysis and Mining of Multialyer Social Networks*. (2019). Available online at: https://cran.r-project.org/package=multinet
42. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing (2019). Available online at: http//wwwR-project.org/
43. Sah P, Leu ST, Cross PC, Hudson PJ, Bansal S. Unraveling the disease consequences and mechanisms of modular structure in animal social networks. *Proc Natl Acad Sci USA.* (2017) 114:4165–70. doi: 10.1073/pnas.1613616114
44. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela JP. Community structure in time-dependent, multiscale, and multiplex networks. *Science.* (2010) 328:876–8. doi: 10.1126/science.1184819
45. Bazzi M, Porter MA, Williams S, Mcdonald M, Fenn DJ, Howison SAMD. Community detection in temporal multilayer networks, with an application to correlation networks. *Multiscale Model Simul.* (2016) 14:1–41. doi: 10.1137/15M1009615
46. Bohlin L, Edler D, Lancichinetti A, Rosvall M. Community detection and visualization of networks with the map equation framework. In: Ding Y, Rousseau R, Wolfram D, editors. *Measuring Scholarly Impact*. Cham: Springer (2014). p. 3–34.
47. Keeling MJ, Rohani P. *Modeling Infectious Diseases in Humans and Animals*. Princeton, NJ: Princeton University Press (2008).
48. Christley RM, Pinchbeck GL, Bowers RG, Clancy D, French NP, Bennett R, et al. Infection in social networks: using network analysis to identify high-risk individuals. *Am J Epidemiol.* (2005) 162:1024–31. doi: 10.1093/aje/kwi308
49. Kao R, Danon L, Green D, Kiss I. Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. *Proc R Soc B Biol Sci.* (2006) 273:1999–2007. doi: 10.1098/rspb.2006.3505
50. VanderWaal K, Enns EA, Picasso C, Alvarez J, Perez A, Fernandez F, et al. Optimal surveillance strategies for bovine tuberculosis in a low-prevalence country. *Sci Rep.* (2017) 7:1–12. doi: 10.1038/s41598-017-04466-2
51. Stella M, Andreazzi CS, Selakovic S, Goudarzi A, Antonioni A. Parasite spreading in spatial ecological multiplex networks. *J Complex Netw.* (2017) 5:486–511. doi: 10.1093/comnet/cnw028
52. Tsai Y-S, Huang C-Y, Wen T-H, Sun C-T. A multilayer epidemiological model integrating daily commuting network. *Sel Top Syst Sci Simul Eng.* (2010)77.
53. Colizza V, Vespignani A. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations. *J Theor Biol.* (2008) 251:450–67. doi: 10.1016/j.jtbi.2007.11.028
54. Saumell-Mendiola A, Serrano MA, Boguñá M. Epidemic spreading on interconnected networks. *Phys Rev E Stat Nonlinear Soft Matter Phys.* (2012) 86:1–5. doi: 10.1103/PhysRevE.86.026106
55. Zheng CY, Xia CY, Guo QT, Dehmer M. Interplay between SIR-based disease spreading and awareness diffusion on multiplex networks. *J Parallel Distrib Comput.* (2018) 115:20–8. doi: 10.1016/j.jpdc.2018.01.001
56. Buono C, Braunstein LA. Immunization strategy for epidemic spreading on multilayer networks. *EPL.* (2015) 109. doi: 10.1209/0295-5075/109/26001
57. Gilbert M, Mitchell A, Bourn D, Mawdsley J, Clifton-Hadley R, Wint W. Cattle movements and bovine tuberculosis in Great Britain. *Nature.* (2005) 435:491–6. doi: 10.1038/nature03548
58. Bates TW, Thurmond MC, Carpenter TE. Direct and indirect contact rates among beef, dairy, goat, sheep, and swine herds in three California counties, with reference to control of potential foot-and-mouth disease transmission. *Am J Vet Res.* (2001) 62:1121–9.
59. Nöremark M, Frössling J, Lewerin SS. A survey of visitors on swedish livestock farms with reference to the spread of animal diseases. *BMC Vet Res.* (2013) 9:184. doi: 10.1186/1746-6148-9-184
60. Sanson RL, Struthers G, King P, Weston JF, Morris RS. The potential extent of transmission of foot-and-mouth disease: a study of the movement of animals and materials in southland, new zealand. *N Z Vet J.* (1993) 41:21–8. doi: 10.1080/00480169.1993.35730
61. Silk MJ, Weber NL, Steward LC, Hodgson DJ, Boots M, Croft DP, et al. Contact networks structured by sex underpin sex-specific epidemiology of infection. *Ecol Lett.* (2018) 21:309–18. doi: 10.1111/ele.12898
62. Azimi-Tafreshi N. Cooperative epidemics on multiplex networks. *Phys Rev E.* (2016) 93:42303. doi: 10.1103/PhysRevE.93.042303
63. Lass S, Hudson PJ, Thakar J, Saric J, Harvill E, Albert R, et al. Generating super-shedders: co-infection increases bacterial load and egg production of a gastrointestinal helminth. *J R Soc Interface.* (2013) 10:20120588. doi: 10.1098/rsif.2012.0588
64. Funk S, Jansen VAA. Interacting epidemics on overlay networks. *Phys Rev E Stat Nonlinear Soft Matter Phys.* (2010) 81:036118. doi: 10.1103/PhysRevE.81.036118
65. Van Reeth K, Braeckmans D, Cox E, Van Borm S, van den Berg T, Goddeeris B, et al. Prior infection with an H1N1 swine influenza virus partially protects pigs against a low pathogenic H5N1 avian influenza virus. *Vaccine.* (2009) 27:6330–9. doi: 10.1016/j.vaccine.2009.03.021
66. Ezenwa VO, Jolles AE. Opposite effects of anthelmintic treatment on microbial infection at individual versus population scales. *Science.* (2015) 347:175–7. doi: 10.1126/science.1261714
67. Wang WH, Chen MM, Min Y, Jin XG. Structural diversity effects of multilayer networks on the threshold of interacting epidemics. *Phys A Stat Mech Appl.* (2016) 443:254–62. doi: 10.1016/j.physa.2015.09.064
68. Stroeymeyt N, Grasse AV, Crespi A, Mersch DP, Cremer S, Keller L. Social network plasticity decreases disease transmission in a eusocial insect. *Science.* (2018) 362:941–5. doi: 10.1126/science.aat4793
69. Neaigus A, Friedman SR, Curtis R, Jarlais DCD, Furst RT, Jose B, et al. The relevance of drug injectors social and risk networks for understanding and preventing HIV-infection. *Soc Sci Med.* (1994) 38:67–78. doi: 10.1016/0277-9536(94)90301-8
70. Balasubramaniam K, Beisner B, Vandeleest J, Atwill E, McCowan B. Social buffering and contact transmission: network connections have beneficial and detrimental effects on Shigella infection risk among captive rhesus macaques. *PeerJ.* (2016) 2016:1–26. doi: 10.7717/peerj.2630
71. Bessell PR, Searle KR, Auty HK, Handel IG, Purse BV, Bronsvoort BMDC. Assessing the potential for Bluetongue virus 8 to spread and vaccination strategies in Scotland. *Sci Rep.* (2016) 6:1–13. doi: 10.1038/srep38940
72. VanderWaal K, Paploski IAD, Makau DN, Corzo CA. Contrasting animal movement and spatial connectivity networks in shaping transmission

pathways of a genetically diverse virus. *Prev Vet Med.* (2020) 178:104977. doi: 10.1016/j.prevetmed.2020.104977
73. Boender GJ, Van Den Hengel R, Van Roermund HJW, Hagenaars TJ. The influence of between-farm distance and farm size on the spread of classical swine fever during the 1997-1998 epidemic in the Netherlands. *PLoS ONE.* (2014) 9:e95278. doi: 10.1371/journal.pone.0095278
74. Machado G, Vilalta C, Recamonde-Mendoza M, Corzo C, Torremorell M, Perez A, et al. Identifying outbreaks of Porcine Epidemic Diarrhea virus through animal movements and spatial neighborhoods. *Sci Rep.* (2019) 9:457. doi: 10.1038/s41598-018-36934-8
75. *OIE-Terrestrial Animal Health Code*. 28th ed. (2019). Available online at: https://www.oie.int/standard-setting/terrestrial-code/access-online/
76. Chaters GL, Johnson PCD, Cleaveland S, Crispell J, De Glanville WA, Doherty T, et al. Analysing livestock network data for infectious disease control: an argument for routine data collection in emerging economies. *Philos Trans R Soc B Biol Sci.* (2019) 374. doi: 10.1098/rstb.2018.0264
77. Davis GH, Crofoot MC, Farine DR. Estimating the robustness and uncertainty of animal social networks using different observational methods. *Anim Behav.* (2018) 141:29–44. doi: 10.1016/j.anbehav.2018.04.012
78. Silk MJ, Jackson AL, Croft DP, Colhoun K, Bearhop S. The consequences of unidentifiable individuals for the analysis of an animal social network. *Anim Behav.* (2015) 104:1–11. doi: 10.1016/j.anbehav.2015.03.005
79. Franks DW, Ruxton GD, James R. Sampling animal association networks with the gambit of the group. *Behav Ecol Sociobiol.* (2010) 64:493–503. doi: 10.1007/s00265-009-0865-8
80. Génois M, Vestergaard CL, Cattuto C, Barrat A. Compensating for population sampling in simulations of epidemic spread on temporal contact networks. *Nat Commun.* (2015) 6:1–13. doi: 10.1038/ncomms9860
81. Kao RR, Green DM, Johnson J, Kiss IZ. Disease dynamics over very different time-scales : foot-and-mouth disease and scrapie on the network of livestock movements in the UK. *J R Soc Interface.* (2007) 4:907–16. doi: 10.1098/rsif.2007.1129
82. Vernon MC, Keeling MJ. Representing the UK's cattle herd as static and dynamic networks. *Proc R Soc B Biol Sci.* (2009) 276:469–76. doi: 10.1098/rspb.2008.1009
83. Konschake M, Lentz HHK, Conraths FJ, Hövel P, Selhorst T. On the Robustness of In- and Out-Components in a Temporal Network. *PLoS ONE.* (2013) 8:e55223. doi: 10.1371/journal.pone.0055223
84. VanderWaal KL, Atwill ER, Isbell LA, McCowan B. Linking social and pathogen transmission networks using microbial genetics in giraffe (*Giraffa camelopardalis*). *J Anim Ecol.* (2014) 83:406–14. doi: 10.1111/1365-2656.12137
85. Blyton MDJ, Banks SC, Peakall R, Lindenmayer DB, Gordon DM. Not all types of host contacts are equal when it comes to *E. coli* transmission. *Ecol Lett.* (2014) 17:970–8. doi: 10.1111/ele.12300
86. Rimbach R, Bisanzio D, Galvis N, Link A, Di Fiore A, Gillespie TR. Brown spider monkeys (*Ateles hybridus*): a model for differentiating the role of social networks and physical contact on parasite transmission dynamics. *Philos Trans R Soc B Biol Sci.* (2015) 370. doi: 10.1098/rstb.2014.0110
87. VanderWaal KL, Picasso C, Enns EA, Craft ME, Alvarez J, Fernandez F, et al. Network analysis of cattle movements in Uruguay: quantifying heterogeneity for risk-based disease surveillance and control. *Prev Vet Med.* (2016) 123:12–22. doi: 10.1016/j.prevetmed.2015.12.003
88. Schulz J, Boklund A, Halasa THB, Toft N, Lentz HHK. Network analysis of pig movements: Loyalty patterns and contact chains of different holding types in Denmark. *PLoS ONE.* (2017) 12:e0179915. doi: 10.1371/journal.pone.0179915
89. Kinsley AC, Perez AM, Craft ME, Vanderwaal KL. Characterization of swine movements in the United States and implications for disease control. *Prev Vet Med.* (2019) 164:1–9. doi: 10.1016/j.prevetmed.2019.01.001
90. Brennan ML, Kemp R, Christley RM. Direct and indirect contacts between cattle farms in north-west England. *Prev Vet Med.* (2008) 84:242–60. doi: 10.1016/j.prevetmed.2007.12.009
91. Cozzo E, Baños RA, Meloni S, Moreno Y. Contact-based social contagion in multiplex networks. *Phys Rev E Stat Nonlinear Soft Matter Phys.* (2013) 88:1–5. doi: 10.1103/PhysRevE.88.050801
92. Liu QH, Xiong XY, Zhang Q, Perra N. Epidemic spreading on time-varying multiplex networks. *Phys Rev E.* (2018) 98:062303. doi: 10.1103/PhysRevE.98.062303
93. Volkova VV, Howey R, Savill NJ, Woolhouse MEJ. Potential for transmission of infections in networks of cattle farms. *Epidemics.* (2010) 2:116–22. doi: 10.1016/j.epidem.2010.05.004
94. De Domenico M, Nicosia V, Arenas A, Latora V. Structural reducibility of multilayer networks. *Nat Commun.* (2015) 6:1–9. doi: 10.1038/ncomms7864
95. Python Software Foundation (2019). Available online at: www.python.org
96. Chen S, White BJ, Sanderson MW, Amrine DE, Ilany A, Lanzas C. Highly dynamic animal contact network and implications on disease transmission. *Sci Rep.* (2014) 4:1–7. doi: 10.1038/srep04472
97. Holme P, Saramäki J. Temporal networks. *Phys Rep.* (2012) 519:97–125. doi: 10.1016/j.physrep.2012.03.001
98. Hassell JM, Begon M, Ward MJ, Fèvre EM. Urbanization and disease emergence: dynamics at the wildlife–livestock–human interface. *Trends Ecol Evol.* (2017) 32:55–67. doi: 10.1016/j.tree.2016.09.012

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

METHODS
published: 17 September 2020
doi: 10.3389/fvets.2020.00604

Autoregressive Models Applied to Time-Series Data in Veterinary Science

Michael P. Ward[1]*, *Rachel M. Iglesias*[2] and *Victoria J. Brookes*[3,4]

[1] Sydney School of Veterinary Science, The University of Sydney, Sydney, NSW, Australia, [2] Australian Government Department of Agriculture, Water and the Environment, Canberra, ACT, Australia, [3] School of Animal and Veterinary Sciences, Faculty of Science, Charles Sturt University, Wagga Wagga, NSW, Australia, [4] Graham Centre for Agricultural Innovation, NSW Department of Primary Industries, Charles Sturt University, Wagga Wagga, NSW, Australia

A time-series is any set of *N* time-ordered observations of a process. In veterinary epidemiology, our focus is generally on disease occurrence (the "process") over time, but animal production, welfare or other traits might also be of interest. A common source of time-series datasets are animal disease monitoring and surveillance systems. Here, we scan the application of methods to analyse time-series data in the peer-reviewed, published literature. Based on this literature scan we focus on autocorrelation and illustrate the recommended steps using ARIMA (Autoregressive Integrated Moving Average Models) methods via analysis of a time-series of canine parvovirus (CPV) events in a pet dog population in Australia, 2009 to 2015. We conclude by identifying the barriers to the application of ARIMA methods in veterinary epidemiology and suggest some possible solutions. In the literature scan the selected 37 studies focused mostly on infectious and parasitic diseases, predominantly for analytical, rather than descriptive or predictive, purposes. Trends and seasonality were investigated, and autocorrelation analyzed, in most studies, most commonly using R software. An approach to analyzing autocorrelation using ARIMA methods was then illustrated using a time-series (week and month units) of CPV events in a pet dog population in Australia, reported to a national companion animal disease surveillance system. This time-series was derived by summing veterinarian reports of confirmed CPV diagnoses. We present data analysis output generated via the R statistical environment, and make this code available for the reader to apply to this or other time-series datasets. We also illustrate prediction of CPV events by rainfall as a covariate. Time-series analysis using ARIMA methods to understand and explore autocorrelation appears to be relatively uncommon in veterinary epidemiology. Some of the reasons might include limited availability of data of sufficient time unit length, lack of familiarity with analytical methods and available software, and how to best use the information generated. We recommend that wherever feasible, such time-series data be made available both for analysis and for methods development.

Keywords: time-series analysis, veterinary science, methods, animal disease, canine parvovirus

OPEN ACCESS

Edited by:
Ioannis Magouras,
City University of Hong Kong,
Hong Kong

Reviewed by:
Fernanda Dorea,
National Veterinary Institute, Sweden
Flavie Vial,
Animal and Plant Health Agency,
United Kingdom

***Correspondence:**
Michael P. Ward
michael.ward@sydney.edu.au

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *31 October 2019*
Accepted: *28 July 2020*
Published: *17 September 2020*

Citation:
Ward MP, Iglesias RM and Brookes VJ (2020) Autoregressive Models Applied to Time-Series Data in Veterinary Science. Front. Vet. Sci. 7:604. doi: 10.3389/fvets.2020.00604

INTRODUCTION

A time-series is any set of N time-ordered observations of a process (1). Within the discipline of epidemiology, our goal is often to understand the underlying processes that generate time-series of disease events. These processes can be explored as part of a time-series analysis, particularly when potential explanatory variables are included as covariates. This can provide insights into disease causation, and thus contribute to the formulation of disease prevention and control programs. However, time-series analysis can also be predictive, with or without covariates. This facilitates the development of forecasting systems to anticipate disease occurrence or detect changes in disease occurrence. Here, we focus on the former goal of understanding disease occurrence.

A key property of time-series is non-independence of values at consecutive time periods. This results in a statistical relationship between values at consecutive time periods and sometimes at different time lags, known as *autocorrelation*. Temporal autocorrelation is a fundamental characteristic of observations recorded over extended periods of time. We can appreciate that daily rainfall data, for example, recorded over a period of months will show autocorrelation: if it rains on a specific day, it is more likely to rain the following day. In addition, rainfall might be more common during certain months, or seasons. Perhaps less obvious is autocorrelation in time-series of disease occurrence. Diseases can be clustered in time due to causes that are autocorrelated (such as climate), due to the methods used to detect disease and the surveillance programs used (for example, certain diagnostic tests only being performed on Mondays, or inspectors at abattoirs working fixed 6-day shifts), and (for infectious diseases) because the number of infected individuals at one time period directly affects the number of infected individuals at a subsequent time period due to disease transmission. Rather than searching for evidence of temporal clustering (2), autocorrelation methods assumes it is present and seek it describe and understand it. Whilst temporal autocorrelation might be expected, often it is subtle.

Autocorrelation makes common statistical approaches inappropriate, and alternative techniques are needed. Time-series analysis invariably begins with descriptive analyses of the dataset under consideration. This consists of separating out ("decomposing") the time-scale dependent characteristics which make up the observed temporal pattern of disease or event occurrence. Broadly, these patterns are the long term (secular), periodic cyclical (if time-independent), and seasonal trends. The aim of this analysis is to characterize temporal patterns. There are a variety of methods for decomposition, including decomposition based on locally-weighted scatterplot smoothing ["seasonal and trend decomposition using locally weight scatterplot smoothing (loess)," STL]; we demonstrate this method in the context of the CPV events. The process of decomposition, whilst attempting to remove autocorrelation from a time-series, also allows an understanding of the autocorrelation itself and its potential causes.

As part of the process of exploring a time-series, *autoregressive models* can be used to determine how much of the observed time-series can be explained by previous observations in the time-series itself. Characterization of temporal patterns—such as trend and seasonality—can be used to understand potential causes of disease. Autoregressive models to describe the occurrence of events based on prior observations include simple autoregressive (AR) models, autoregressive moving average (ARMA) and autoregressive integrative moving average (ARIMA) models, which differ in the way previous values in the time-series are used to describe future values. AR models are essentially linear regressive models for which each regression term is a time-lagged value (i.e., a value measured at a previous time point—the "lag") of the same time-series. MA models instead use lagged values of forecast errors, and ARMA models combine both. ARIMA models can also include differencing (i.e., the value at one time point is subtracted from the value at another time point) of the series. Causation can be further investigated by multivariate models. For example, autoregressive models can be extended to include covariates, and in a further extension, information from more than one time-series can be used in vector autoregressive models to forecast future values of each time-series. We demonstrate the way in which visual exploration of autocorrelation function (ACF) and partial autocorrelation function (PACF) plots can provide insights into how to fit a model, and how to select the best model fit for ARMA and ARIMA models.

We begin our discussion of the analysis of time-series data in veterinary epidemiology from our perspective that ARIMA methods are not commonly applied within the discipline. In situations in which methods to analyse time-series data have been applied, we investigate the more commonly used methods and data sources reported via a scan of recent literature. This is motivated by an appraisal of current usage and gaps in the field, rather than a comprehensive, systematic review, to provide the reader with a range of literature in which methods for analysis of times-series data have been used. We then demonstrate the application of autoregressive models using ARIMA methods on a surveillance dataset, and make recommendations to increase the use of such methods in veterinary science.

LITERATURE SCAN

CAB Abstracts Index via Web of Science was searched using TOPIC: (time-series) and TOPIC: (analysis) and TOPIC: (veterinary) during the timespan 1980 to present (31 August 2019), restricted to English language journal articles only. The titles of all articles returned by this search were screened for scope [time-series analysis methods applied to animal (including zoonotic) diseases]. Note that studies in which *time-series data* were reported, but which did not describe the application of *time-series analysis* methods, were excluded.

A template was developed—via discussion between the authors—to extract information from each article (see **Supplementary Table 1**). Full versions of the subset of articles

were then obtained and randomly assigned to one of the three authors.

In total, 60 articles (see **Supplementary Table 2**) were identified. Of these, five were unavailable for review and 18 were out-of-scope. The latter included articles in which the primary event was a disease in humans only (for example, dengue fever, Crimean-Congo haemorrhagic fever, tick-borne encephalitis, Ross River fever), or the focus was on detection of aberrations within a time-series [for example, (3)]. We applied these exclusions because our aim is to introduce readers to autoregressive models and applications to animal diseases.

Of the remaining 37 articles, publication year ranged from 1990 to 2019 and studies were conducted in 19 different countries (**Supplementary Table 2**). One study was conducted at the global scale [highly pathogenic avian influenza; (4)]. Data used in these studies were derived from surveillance systems (including internet searches) (14); monitoring systems (11), for example slaughterhouse recording systems; clinical records (6); laboratory records (3); and bespoke research projects (3). These studies were focused mostly on livestock (26). The temporal unit of data collection was most commonly day (17) or month (16), and the median period (years) covered by the datasets analyzed was 10 (IQR 5–16).

The studies identified focused on a wide range of events, but mostly either specific infectious diseases (e.g., rabies) or defined syndromes (e.g., pleurisy and pneumonia).

The purpose of the time-series analysis performed was either analysis (18), description (12), or prediction (7). Studies were considered descriptive if they included only visualization of the time-series or descriptive statistics, whereas those that also included decomposition of the series, or developed models of the time-series, were considered analytical. Those that used the models to predict trends beyond the range of the time-series were considered predictive. Data analyzed was most commonly counts of events. Where data was manipulated before analysis, aggregation to a coarser temporal unit was most common.

Analysis of trends was performed in most (27) studies, mainly using regression models (13). Autoregression was analyzed in the majority of studies (23). In six of these, autocorrelation and partial autocorrelation functions (ACF and PACF; see section An Example of Time-Series Analysis Methods—Canine Parvovirus Reports for Definitions and Methods) were used, and in other studies (10) modeling approaches were used, including autoregressive models. ARMA or ARIMA models were described in 13 of the 23 studies in which autoregression was analyzed. Seasonality was analyzed in 28 studies, however the methods used varied greatly; for example, visual, ACF and PACF, seasonal autoregressive models, automated exponential smoothing state space models, periodograms, and seasonal and trend decomposition STL.

Forecasting was undertaken in 12 studies. The most common (18) software used to analyse time-series data was R.

We observed that the most often cited advantage of using time-series analysis methods was the ability to predict disease occurrence, contributing to early warning and therefore disease prevention. Some of the barriers discussed include the scarcity of long-term, computerized, automatically collected, and publicly available data; identifying outbreak or disease-free baselines; event data sparseness (excessive zeros); data aggregation (temporal scale); time gaps in the data; lack of constant population at-risk; and model validation.

In summary, in this literature scan, time-series analysis methods in veterinary science were mostly focused on infectious and parasitic diseases, analyzed by decomposing and modeling the time-series. This approach most often involves investigation of trends and seasonality, and analysis of autocorrelation, usually aided by the use of R software. Based on this, we next illustrate methods that can be used to investigate and analyse trends, seasonality and autocorrelation in veterinary science by presenting a step-by-step guide to analysis of a canine parvovirus time-series using R.

We focus on ARIMA methods because beyond a description of the trend and seasonality of time-series data, ARIMA models are an accessible method to describe autocorrelations within data and assess the influence of covariates such as climate variables. These methods can be considered a foundation in autoregressive methods for time-series analysis. Other methods—such as aberration detection algorithms, stochastic modeling approaches and machine-learning methods—can then be investigated for applications requiring long-term prediction (5–7).

AN EXAMPLE OF TIME-SERIES ANALYSIS METHODS—CANINE PARVOVIRUS REPORTS

Prior to embarking on autoregressive modeling, we need to consider when it is appropriate to apply these methods—and when it is not. For such analysis, a dataset of sufficient length and completeness needs to be available. Without sufficient data, it is difficult to identify trends and patterns, to build models, and determine statistical significance. In veterinary science, data generated by monitoring and surveillance systems are often analyzed by autoregressive modeling (see section Literature Scan). However, missing data can be an issue (see section Results of Analyzing a Time-series of Canine Parvovirus Reports), as can data gaps in the time-series caused by temporary interruptions to data collection. Assuming a stable population at-risk simplifies analysis and interpretation of results, but such assumptions need to be plausible. Other more general epidemiological issues—such as selection, ascertainment and measurement bias—also are applicable to autoregressive modeling and need to be considered.

Here we describe an analysis using autoregressive methods as an example that readers can use to guide their own analyses (8). The data and R code used for the analysis are available at https://zenodo.org/record/3738684#.X1HOYNZuLIU (accessed 04/09/2020).

Our time-series analysis begins with a description of the data, including the source, results of initial data checking and any manipulation required to make it suitable for time-series analysis. The time-series is then plotted, and secular and seasonal trends are assessed using decomposition then linear regression. Before fitting an autoregressive model, the series is assessed for stationarity using graphical and statistical methods. Stationarity

is a key requirement to fit models to time-series data. A series is considered stationary if it is not changing systematically over time. A method for inducing stationarity—differencing—is also explained and demonstrated. We then fit a number of ARIMA models and use these to forecast disease cases beyond the range of the dataset. Finally, we investigate the influence of a covariate (rainfall) on the time-series and give a brief example of how cross-correlation and vector autoregressive models can be used to investigate relationships in time-series. We present the analysis of the example dataset in a stepwise guide to assist the reader to replicate the approach on this or other, similar datasets.

We have used the R statistical environment (9) for all analysis described. For readers not familiar with this platform, introductory courses and tutorials are widely available online and we recommend spending some time familiarizing yourself with the program before attempting this analysis. The code provided in the Zenodo repository will work if you have R correctly installed and operating on your computer and have installed the packages listed below.

The following packages for data visualization, manipulation and analysis of time-series data used in this analysis: ggplot2 (10), plyr (11), dplyr (12), lubridate (13), tseries (14), vars (15, 16), and forecast (17, 18).

To align readers to the associated R code, the corresponding "chunk" (C) in the code (https://zenodo.org/record/3738684#.X1HOYNZuLIU, accessed 04/09/2020) is included in the methods below. Chunks C1-C3 initiate and load the required packages.

Here, we present a series of six steps to guide the reader in applying time-series analysis to the example dataset.

Step 1: Describing the Data

This worked example uses data from the Disease Watchdog system, in operation since 2010 in Australia and initiated to collect information on infectious diseases of dogs and cats in Australia (19–22). By 2015, nearly 25,000 disease cases and 19,000 reports had been submitted. The system was deactivated in early 2017.

Veterinarians and veterinary clinic staff were the contributors of data within this system. Besides disease diagnoses and their date of occurrence and postcode of residence, a range of other patient data was also collected, including age, sex, neuter status, breed, diagnostic method, and vaccination status. To encourage timely reporting, data was used to produce near-real time disease maps which veterinarians accessed to educate their clients (19). In this example, canine parvovirus (CPV) is used as the event of interest. CPV is a highly contagious disease of dogs and an important cause of morbidity and mortality in young dogs (23). It has a worldwide distribution and occurs as endemic disease or as local outbreaks.

Records of all CPV cases reported Australia-wide between October 2009 and November 2015 were extracted from the Disease Watchdog database. For analysis, cases which were reported to have been vaccinated at any time were excluded. Furthermore, only those cases in which the diagnosis of parvovirus had been confirmed by diagnostic testing were included. To illustrate approaches to analyzing time-series data, we applied these methods to events only, where an event consists of one or more cases reported by the same veterinarian with the same date of occurrence. We also restricted analysis to events reported from the state of New South Wales.

The dataset was loaded (C4) and checked for duplicated or missing data (C5). The number of events, and minimum and maximum dates of occurrence were reported (C6). The number of parvovirus events were then aggregated by week and by month (based on the reported date of occurrence) to create two time-series datasets (weekly and monthly) for subsequent analyses (C7−9).

Step 2: Visualization

Summary information on CPV events was calculated for the time series at both the weekly and monthly aggregation, and each dataset was plotted with a smoothed curve of events overlaid to visually assess trend (C11). The smoothing process in R is achieved by loess regression (see section Step 4 for a technical explanation of this method). This is exploratory analysis that can be used to inform further analytical approaches. Smoothed curves for both events/week (**Figure 1**) and events/month (**Figure 2**) demonstrate a decreasing trend over time, with the frequency of events being relatively stable during the period 2010 to 2013. If the aim of the analysis was to investigate risk factors for the pattern of events observed, this might suggest that the time-series can be truncated to the period 2010 to 2013, inclusive. If changes in CPV surveillance are of interest, further analysis might include the entire time-series. In addition, these initial plots and smoothed curves can inform the temporal scale of analysis. Visual assessment of **Figures 1** and **2** suggests that monthly aggregation of events is sufficient to preserve the patterns present in the data. However, if the aim of analysis is to identify covariates associated with these patterns, the temporal units used to collect covariate data would also need to be considered.

Step 3: Linear Regression

After conversion of the events series to a computer-recognized "time-series object" (C12), linear regression analysis was used to further explore and quantify secular and seasonal trends (C13). The outcome was the number of events per week (or per month) and the predictors were time in weeks (or months) to assess trend, and week (or month) of the year to assess seasonality. Linear regression is used to confirm impressions from time-series plots and smoothed curves (step 2), to test the statistical significance and to quantify these trends. Identifying such trends is a major component of analysis of time-series data, and can lead to hypothesis-generation regarding potential causes of such patterns.

Step 4: Decomposition

The time-series were then decomposed to separately visualize temporal components including trend and seasonality and the remainder component (also known as "random" or "white noise"). Again, such visualization facilitates the identification and characterization of patterns and potentially what might be

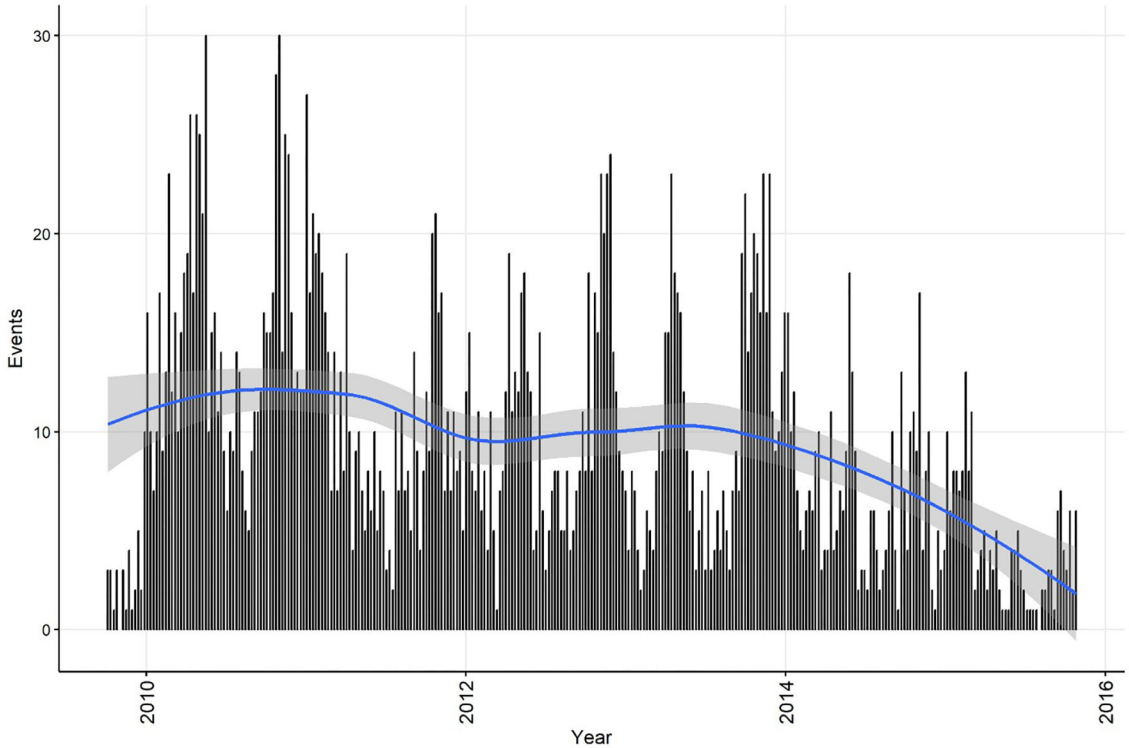


FIGURE 1 | Confirmed canine parvovirus events/week reported from New South Wales in a surveillance system in Australia, 2009–2015. Blue line, loess smoothed curve of events/week with 95% CI (gray).

causing such patterns. For example, the trend or seasonal pattern might dominate. Alternatively, removing trend and season might still result in the remainder time-series showing a discernable pattern. This suggests greater complexity in the time-series (or the incorrect choice of window size to calculate trend and seasonal components).

Two methods were used: moving averages and "seasonal and trend decomposition using loess" (STL; C14). Both are additive models of the form $Y[t] = T[t] + S[t] + e[t]$ in which $Y[t]$ is the model output at time t, $T[t]$ is the trend component at time t (which includes cyclical and longer trend patterns, the "trend-cycle" component), $S[t]$ is the seasonal component at time t and $e[t]$ is the remainder (or residual i.e., what remains in the time-series after removing seasonal and trend components) at time t. If the variance of the trend or seasonal components of the time-series is not constant throughout the time-series, a multiplicative decomposition is likely to be more appropriate than an additive model.

In moving averages the trend component is determined using a moving average window of an appropriate width. This trend component is then subtracted from the original values, and the data grouped by the seasonal element and averaged for each season. The seasonal component is determined by subtracting the average of the seasonal averages from each seasonal average.

A challenge is the choice of an appropriate moving average width. A default width of three time units can be chosen, meaning that for every observation in the time-series, the observation immediately preceding and immediately following that observation is used to calculate an average value. If data within a time-series have been collected with a known periodicity (for example, observation of disease conditions at an abattoir collected every Monday and Tuesday), this could also be used to inform the moving average width.

The STL method is an iterative process that recalculates the seasonal and trend components by a loess smoothing procedure that initially fits a low-order polynomial to the data. A robustness weighting is calculated for each time point between each iteration, and incorporated into the smoothing procedure in the next iteration, which also uses the trend component from the previous iteration (24).

Step 5: Fitting Autoregressive Models

Once the time-series has been explored using the methods above, we use the information gained from these analyses to select and fit an ARIMA model. For demonstration and due to the findings in these exploratory analyses, seasonal autoregressive models with an ARIMA structure were then fitted to the time-series. Autoregression is the relationship between values in a time-series and values in that same time-series measured previously in time (the lag). For example, an autoregressive model of lag 1 describes the relationship between observations and their value in the preceding time unit. The Auto Regressive (AR) terms refer to the number of lagged values in the model. In the non-seasonal part of the model, the order of lagged values is termed "p," and in the seasonal part of the model the order of lagged values is termed

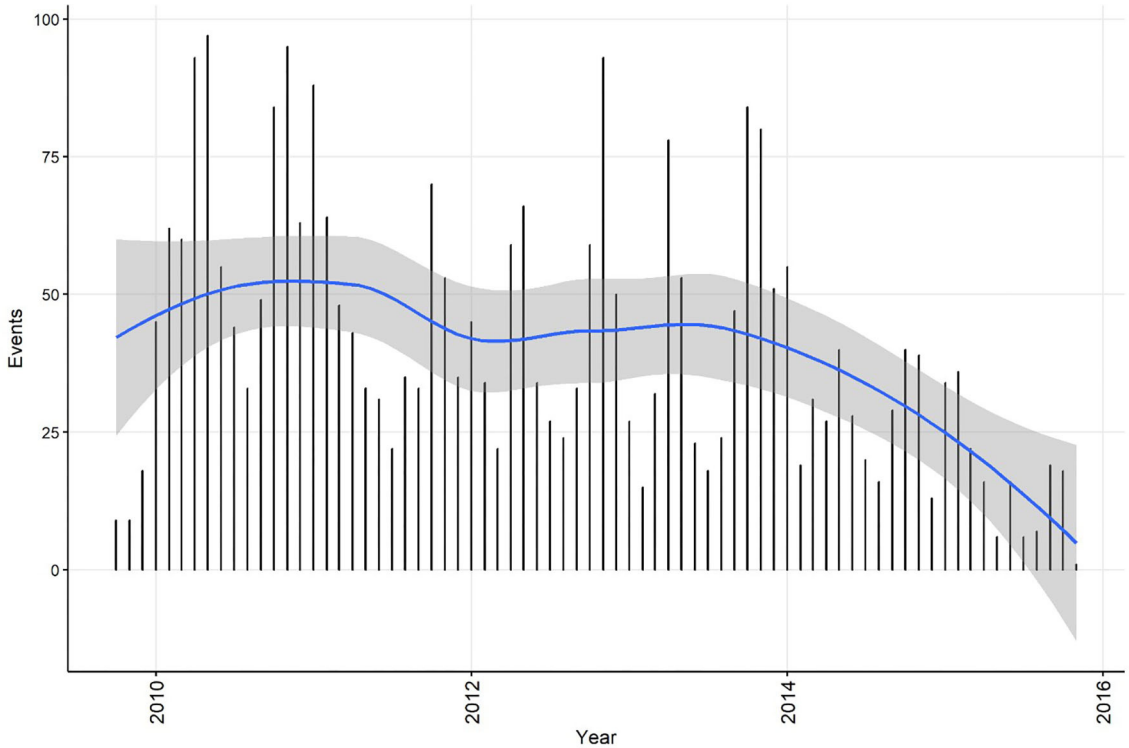


FIGURE 2 | Confirmed canine parvovirus events/month reported from New South Wales in a surveillance system in Australia, 2009–2015. Blue line, loess smoothed curve of events/month with 95% CI (gray).

"*P*." Moving Average (MA) terms—not to be confused with the calculation of a moving average in series decomposition—refer to the number of lagged errors in the model. It is essentially the relationship between current and lagged errors in the time-series. In the non-seasonal part of the model, the order of lagged errors is termed "*q*," and in the seasonal part of the model the order of lagged errors is termed "*Q*." Integration (*I*) terms refer to the number of differences used to make the time-series stationary. In the non-seasonal part of the model, the order of differences is termed "*d*," and in the seasonal part of the model the order of differences is termed "*D*." The overall structure of the model can be written as *(p, d, q) (P, D, Q) m*, in which *m* refers to the number of time-series observations in a seasonal cycle.

The time-series (weekly and monthly reported CPV events) were assessed for stationarity to determine the orders for *d* and *D* to use in the ARIMA model. Initially, an automated function in R was used to determine if differencing was required for both the non-seasonal components (*d*) and seasonal components (*D*) of the ARIMA model using a sequence of unit root tests (KPSS test as default, C16). Stationarity was then further assessed using visualization of time-series plots, auto-correlation function (ACF) plots, and statistical tests (C17−18). Statistical tests included the Ljung-Box test, the Augmented-Dickey Fuller (ADF) test and the Kwiatowski-Phillips-Schmidt (KPSS) test. In the case of a non-stationary time-series, the time-series was first-differenced and assessed again for stationarity. The objective of applying this range of methods is to ensure that any need for differencing—either non-seasonal or seasonal—is identified. Some methods (particularly statistical tests) might not suggest the need for differencing in specific datasets, so a conservative approach is to apply several methods.

ACF and PACF plots were also used to assess the moving average (MA; q, Q) and autoregressive (AR; p, P) non-seasonal and seasonal components of the weekly and monthly ARIMA models following differencing (C19). The ACF plot allows us to visualize the correlation between values in the series and values lagged at a certain number of time points previously, whereas the PACF plot shows the correlation between values in the series and those at a given lag after removing the effect of values at intervening lags. ACF plots can indicate the moving average order *q* to include in an ARIMA model i.e., the lag at which autocorrelation becomes statistically non-significant. Similarly, the PACF plot can inform on the autoregressive order *p* to include. These functions can also be used to inform on seasonal moving average and autoregressive orders, respectively. We give a practical demonstration of how to interpret ACF and PACF for the purposes of ARIMA model parameterization in section An Example of Time-Series Analysis Methods—Canine Parvovirus Reports, using the time-series of CPV events.

Auto-fitting was used to select a starting model (C20−21). Further models were constructed that were simpler (lower parameter terms than the auto-fitted models) but still within the parameter terms for *(p, d, q) (P, D, Q)* that were estimated during exploratory analysis (C22). The models with the lowest

Akaike Information Criterion (AIC) estimates were selected. Model fit was assessed by visualization of predicted time-series relative to observed time-series, and examination of residuals for stationarity (time-series plot, ACF plot, Ljung-Box test) and normality. Because of the auto-fitting algorithms used to identify candidate ARIMA models, it is important to also visualize model(s) selected to ensure these make logical sense and have a biological explanation. Once a final model has been selected, it can be used to predict events for a specified time period beyond the range of the time-series. A predictive model can form the basis of a forecasting system, in which timely anticipation of disease events allows response strategies to be implemented. There are examples of forecasting in veterinary science using time-series analysis (see section Literature Scan). We demonstrate the use and interpretation of these methods in the context of the CPV data below.

Step 6: Multivariate Analysis

To illustrate multivariate time-series analysis methods, a corresponding time-series of rainfall was created. The center of the postcodes in NSW from which CPV was reported during the study period was identified. This was achieved by joining case and event data to a polygon shapefile of NSW postcodes (ArcGIS v. 10.5. ESRI). We then identified the central feature (Spatial Analyst. ESRI), postcode 2850. From this postcode, a Bureau of Meteorology weather recording station was identified [Mudgee (062021), 32.58°S, 149.58°E] and daily rainfall data during the period 1 January 2009 to 31 December 2015 was extracted [1]. Any missing data in the time-series were supplemented by accessing data from the closest weather recording station [Mudgee Airport AWS (062101)]. The rainfall time-series was then aggregated to a monthly time unit to produce a time-series of total monthly rainfall. Dependent on the data, other metrics might be more appropriate, such as monthly median daily temperature or total monthly degree-days.

Covariate time-series datasets are often derived secondarily to the primary time-series of interest (often disease data in veterinary science). Besides climate (including rainfall, temperature and humidity), time-series data might be available on economic indicators, landscape and environmental variables and demographics. For analysis, data need to have the same temporal scale and duration (including time lags) as the primary time-series of interest, and should also broadly match the spatial extent (i.e., when covariates are used, they should be derived from the same area as the outcome of interest, rather than from a larger or a different area).

The presence of substantial data gaps in the series (other than randomly distributed missing data as in our CPV–rainfall example) can render such series unusable if it is not possible to impute data.

The rainfall data were prepared, described and decomposed to assess temporal trends (C23). Quantitative assessments further investigated the trend, seasonality and need for differencing (C24). An automated function was used to fit a dynamic model (ARIMA with rainfall as a predictor) to the CPV and rainfall time-series (C25). Model fit was assessed by visualization of the predicted time-series relative to the observed time-series, and examination of residuals for stationarity (time-series plot, ACF plot, Ljung-Box test) and normality.

Finally, a vector autoregessive model was fit to the CPV and rainfall time-series following examination of a cross-correlation plot between the CPV and rainfall time-series (C26−28). These models assume that a bi-directional relationship ("feedback") between the variables is possible. Whilst this might be a useful premise in the context of time-series of disease in different populations (for example, "who infects whom?"), in the context of this dataset this is implausible (CPV events cannot cause rainfall). However, we include the code for demonstration purposes.

RESULTS OF ANALYZING A TIME-SERIES OF CANINE PARVOVIRUS REPORTS

Step 1: Data Description

Between 2009 and 2015, a total of 24,602 cases and 19,048 events were reported in the Disease Watchdog system. Of these, 20,182 and 15,499 respectively were dog cases and events. During this time period, there were a total of 7,933 CPV cases and 5,837 CPV events reported.

Following application of selection criteria (diagnostic method, nil vaccination history), a total of 2,987 events (3,584 cases) remained for analysis (1.2 cases per event). The earliest and latest reporting dates were 6 October 2009 and 1 November 2015, respectively. The duration of the time-series dataset was 2,218 days, 315 complete weeks and 74 complete months. The median (range) number of cases reported per week was 9 (1–45), and the median (range) number of events reported per week was 8 (1–30).

Step 2: Visualization

The temporal distributions of weekly and monthly events are shown in **Figures 1**,**2**, respectively. A decrease in reported events during the period was apparent in both time-series (indicated by the blue line generated by a loess smoothing function). There were no gaps in the time-series of events.

Step 3: Linear Regression

Linear regression analysis indicated that the decrease in events/week was 1.22 each year (95% CI 0.87−1.57 events/week each year). Statistically significant ($P \leq 0.05$) decreases in the number of events were observed in weeks 11, 24, 26−29, 31−35, and 38. Linear regression analysis indicated that the decrease was 5.76 monthly events/year (95% CI 3.02−8.50 events/year). Statistically significant ($P \leq 0.1$) decreases in the number of events were observed in July and August (winter season).

Step 4: Decomposition

Plots of the decomposed time-series are shown in **Figures 3**,**4**. The trend lines were consistent with **Figure 2**, and seasonal cycles were apparent. The weekly and monthly seasonal cycles were overlaid in **Figure 5** and illustrated that whilst the patterns were consistent with the regression analyses and with each other, monthly seasonality had a simpler, less variable pattern

[1] http://www.bom.gov.au/climate/data/ accessed 30 September 2019.

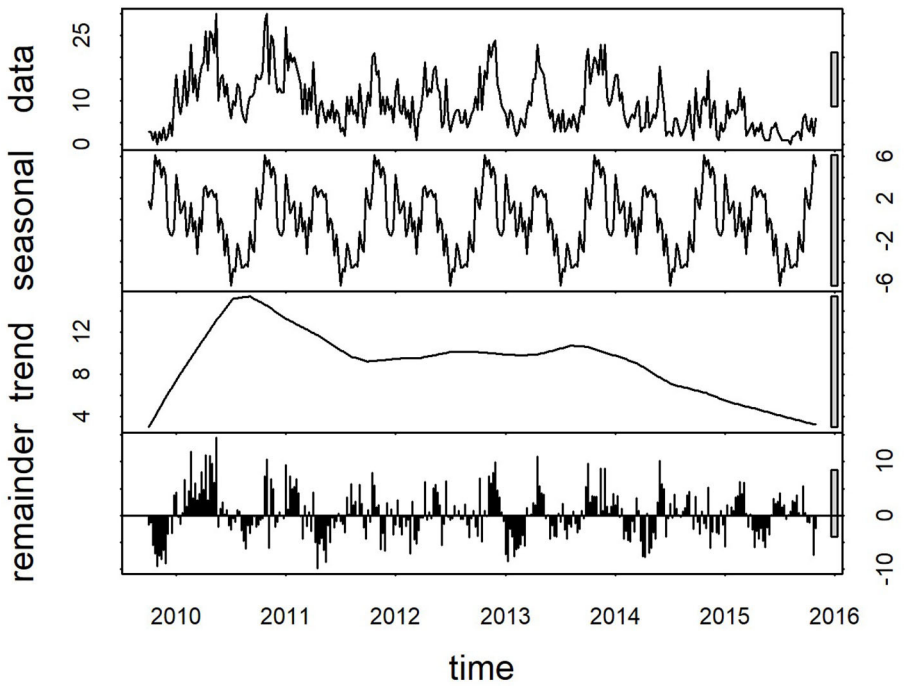


FIGURE 3 | Components of a time-series of weekly confirmed canine parvovirus events reported from New South Wales in a surveillance system in Australia, 2009–2015. Y-axis units, parvovirus events/week. Gray bars on right y-axis indicate the equivalent magnitude of variation of each component ("trend," "seasonal," "remainder") relative to the "data" series, which demonstrates that most variation in the series is in the "remainder" component.

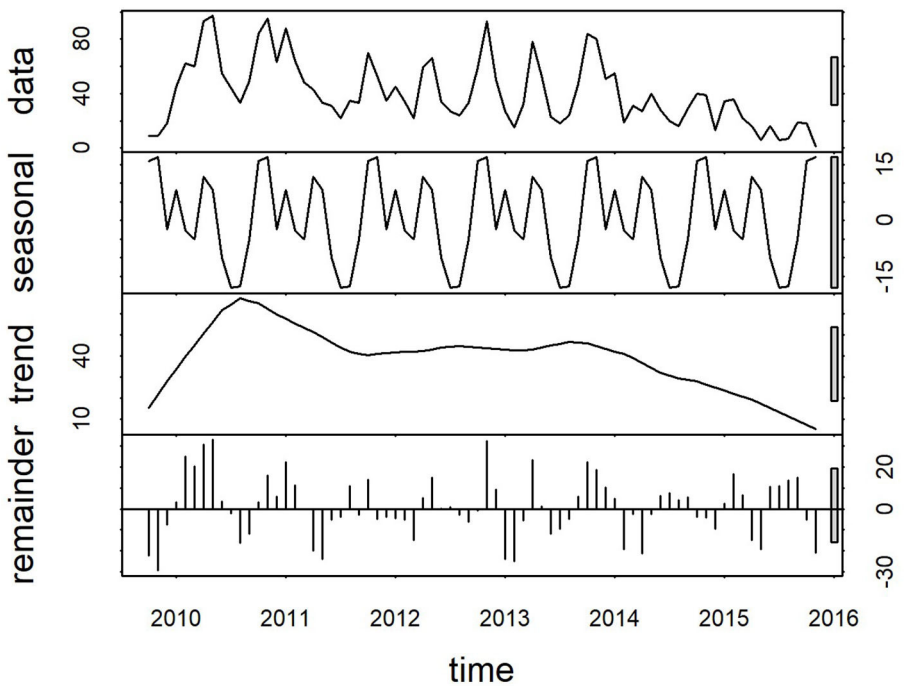


FIGURE 4 | Components of a time-series of monthly confirmed canine parvovirus events reported from New South Wales in a surveillance system in Australia, 2009–2015. Y-axis units, parvovirus events/month. Gray bars on right y-axis indicate the equivalent magnitude of variation of each component ("trend," "seasonal," "remainder") relative to the "data" series, which demonstrates that most variation in the series is in the "remainder", and "trend" components.

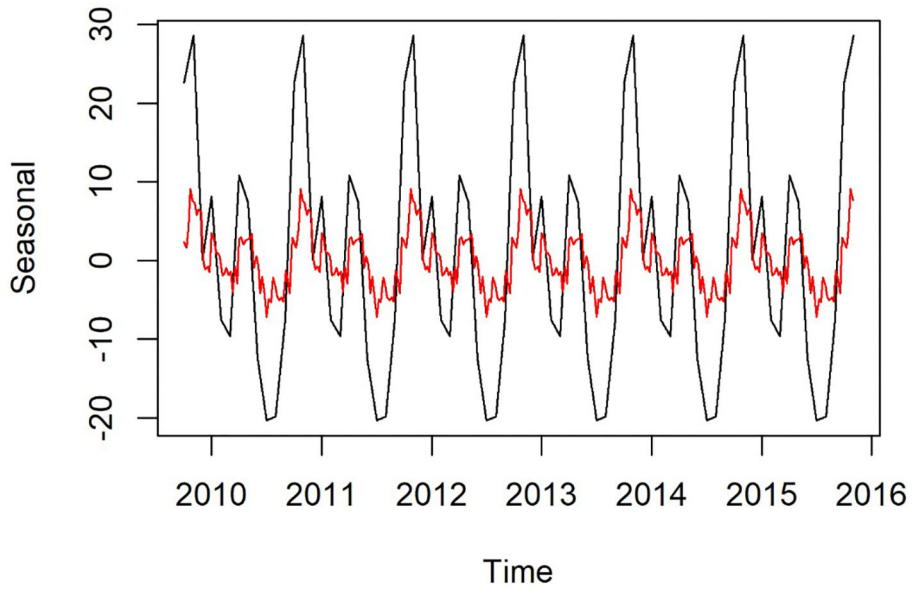


FIGURE 5 | Seasonal component of time-series of weekly **(red)** and monthly **(black)** confirmed canine parvovirus events reported from New South Wales in a surveillance system in Australia, 2009–2015.

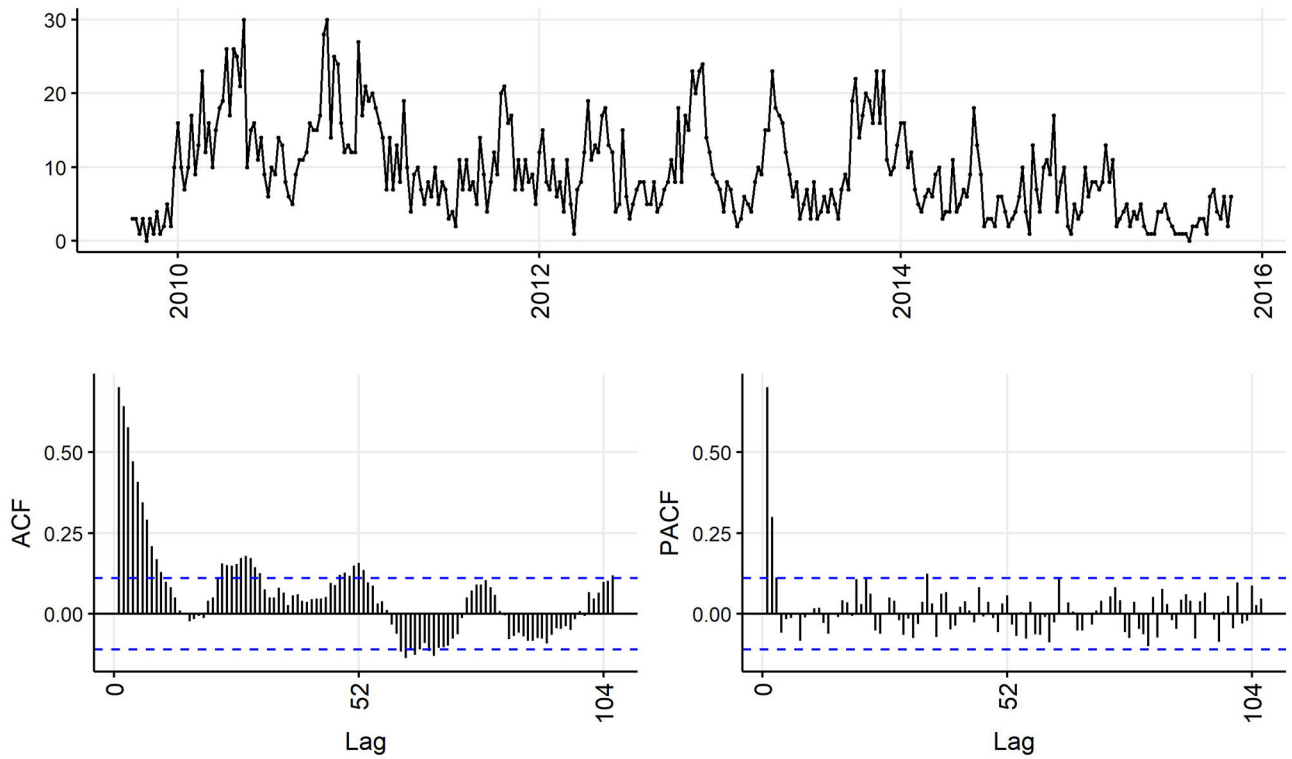


FIGURE 6 | Time-series **(top)**, autocorrelation function **(left)**, and partial autocorrelation function **(right)** plots of weekly aggregated confirmed canine parvovirus events reported in a surveillance system in Australia, 2009–2015.

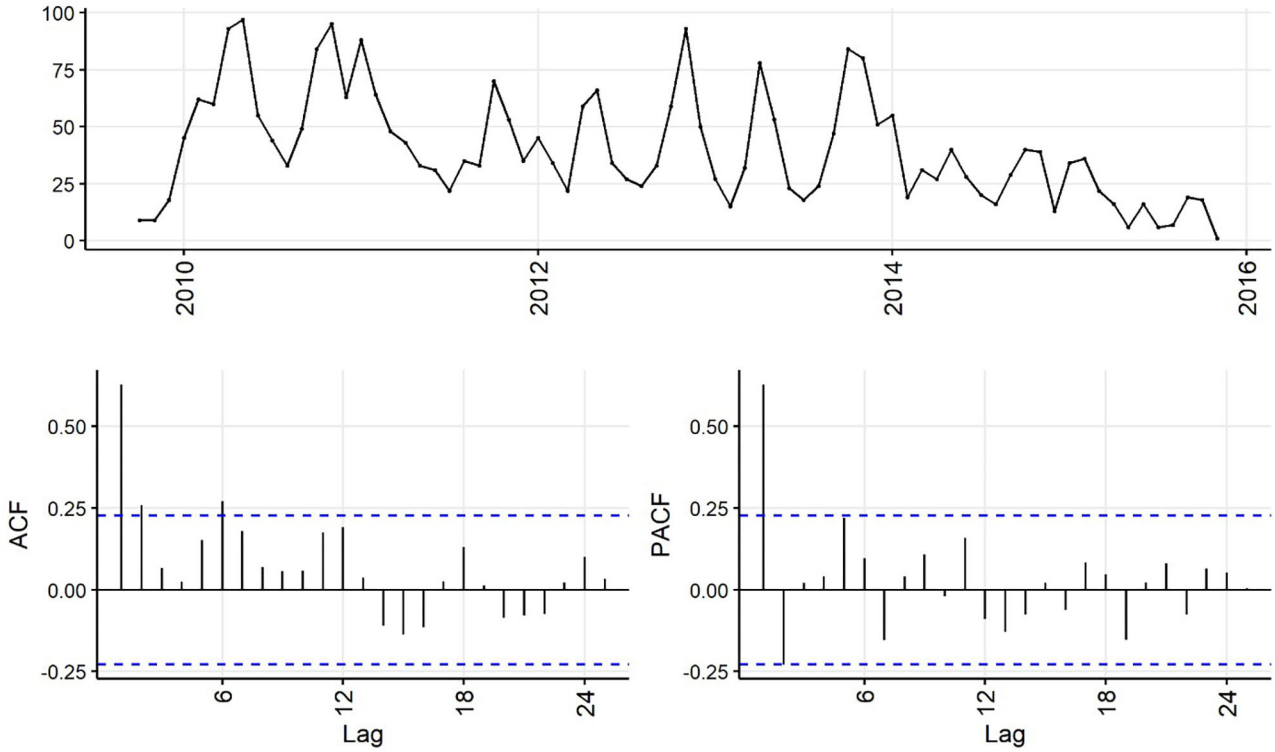


FIGURE 7 | Time-series **(top)**, autocorrelation function **(left)**, and partial autocorrelation function **(right)** plots of monthly aggregated confirmed canine parvovirus events reported in a surveillance system in Australia, 2009–2015.

than weekly seasonality. Both weekly and monthly remainder components appeared to oscillate symmetrically around zero.

Step 5: Autoregressive Models

Decreasing trend in the weekly and monthly time-series, as well as significant autocorrelation of 10 weeks and 2 months in weekly and monthly ACF plots, respectively, suggested non-stationarity (**Figures 6**,**7**). Automated testing of both weekly and monthly series with a sequence of KPSS tests suggested that first differencing of one order would make the series stationary for the non-seasonal component of subsequent ARIMA models, and that differencing was not necessary for the seasonal components of these models. The differenced time-series plots and ACF plots of the weekly and monthly time-series were plausibly stationary—trend was less apparent and there was only one lag of significant autocorrelation in the weekly ACF plot and no initially autocorrelated lags in the monthly ACF plot (**Figures 8**,**9**).

Statistical tests of the weekly and monthly raw time-series were consistent with these findings (**Table 1**). Ljung-Box tests of both time-series suggested non-independence ($P < 0.05$). This was expected due to the autocorrelation observed in the ACF plots. ADF tests of both series suggested stationarity around trend ($P > 0.05$). This was consistent with the time-series plots in which there was decreasing trend but symmetrical oscillation of the series around this trend. KPSS tests for trend-stationarity also suggested trend-stationarity for both time-series ($P > 0.1$), and lack of level-stationarity (trend was present) for both series ($P = 0.01$).

Statistical test results for stationarity of the differenced time-series were similar, except that the KPSS test suggested level-stationarity (no trend) for both series. Overall, given the observed series, ACF and PACF plots and the findings of the statistical tests, both differenced time-series appeared more stationary than raw weekly and monthly time-series, suggesting $d = 1$ of the non-seasonal part of the weekly and monthly ARIMA models.

The ACF plot of the weekly time-series has a fast initial decay with only the first lag significant. This indicates $q = 1$ for the weekly ARIMA model. The ACF plot of the monthly time-series has limited autocorrelation at 2 lags. This could indicate $q = 0{-}2$ for the monthly ARIMA model. The PACF plot of the weekly data has a fast decay with significant partial autocorrelation in the first two lags. This suggests $p = 2$. The PACF plot for the monthly data has limited partial autocorrelation significant. This suggests $p = 0{-}2$ for the monthly ARIMA model.

For seasonality, there are spikes in the weekly ACF at approximately 2 years, indicating $Q = 1{-}2$. There are 3 spikes around 6 months in the PACF, indicating $P = 3$. For seasonality in the monthly data, there are consistent spikes at 6 months, suggesting $Q = 2$, and limited spikes in the PACF, suggesting $P = 0{-}1$.

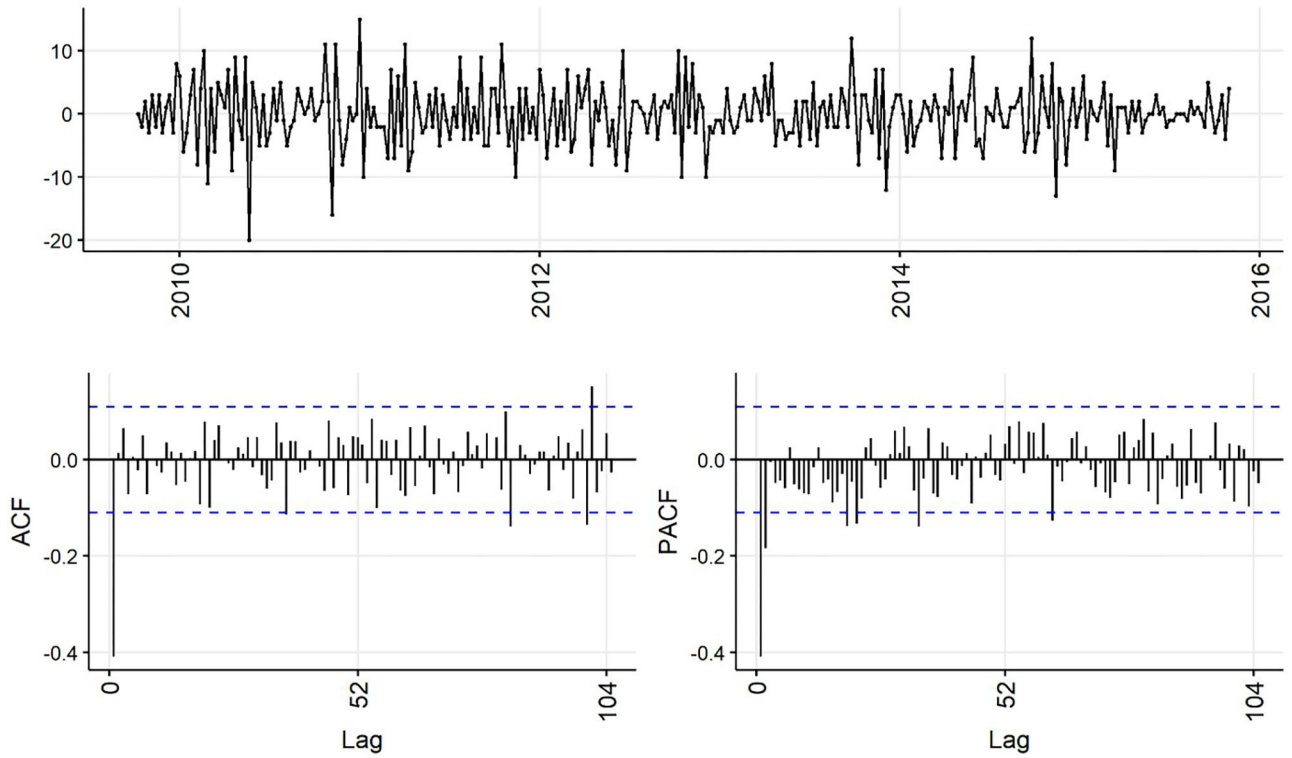


FIGURE 8 | Time-series **(top)**, autocorrelation function **(left)**, and partial autocorrelation function **(right)** plots of differenced weekly aggregated confirmed canine parvovirus events reported in a surveillance system in Australia, 2009–2015.

Auto-fitted ARIMA models for weekly and monthly time-series had (p, d, q) (P, D, Q) [m] structures (3,1,1) (1,0,0) [52] with drift to allow a decreasing trend over time (AICc = 1834.24) and (2,1,1) (2,0,0) [12] (AICc = 632.61). Other simpler structures were assessed with reduced orders for *(p, d, q) (P, D, Q)* that were still within the orders estimated in the exploratory analysis.

The final selected ARIMA models for the weekly and monthly time-series had the structures (2,1,1) (1,0,0) [52] with drift and (2,1,1) (2,0,0) [12] (auto-fitted model), respectively. Parameter values are shown in **Tables 2**,**3**. These models had the simplest structure and lowest AICc, plausible forecast plots, and reasonably normally distributed residuals that were time-independent (ACF plots of residuals and Ljung–Box test; $P > 0.05$).

Step 6: Multivariate Analysis

Although decomposition of the monthly rainfall time-series (**Figure 10**) suggested a decreasing trend and seasonality, neither were quantitatively significant. The rainfall time-series appeared stationary (visually as a time-series and with ACF and PACF plots, and also following statistical tests). Automated model fitting of an ARIMA model of the monthly CPV events time-series with rainfall as a predictor suggested that there was an association between the previous 3 months' rainfall and CPV events (**Table 4**). The coefficients indicate that current and prior rainfall are associated with an increase in parvovirus cases—whilst current and recent (1–2 month lags) rainfall are associated with an increase in cases currently, rainfall 3 months previously are associated with a reduction in cases reported. Although all but the 1 month rainfall lag coefficient are not statistically significant (95% CIs include 1; $P > 0.05$), all variables are required in this model to produce the best model fit.

Interpretation

Without detailed interpretation of the epidemiology of CPV in NSW, general interpretation of the output from the above analysis and some observations are as follows. During the study period, most cases of CPV occurred as individual case reports rather than events, but focusing on events (in which cases are likely epidemiologically-linked) produces information that is more meaningful for disease control and prevention. The selection criteria applied mean that this event time-series is accurate, even though it might not represent the entire study population (owned dogs in NSW between 2009 and 2015) because of the voluntary nature of reporting within the surveillance system. The length of the time-series analyzed is 2,218 days. Although this is a large size (*N*), daily fluctuations in reporting necessitate aggregation to the week and month level to better understand trends and patterns. In addition, knowledge of daily patterns of occurrence and reporting are unlikely to

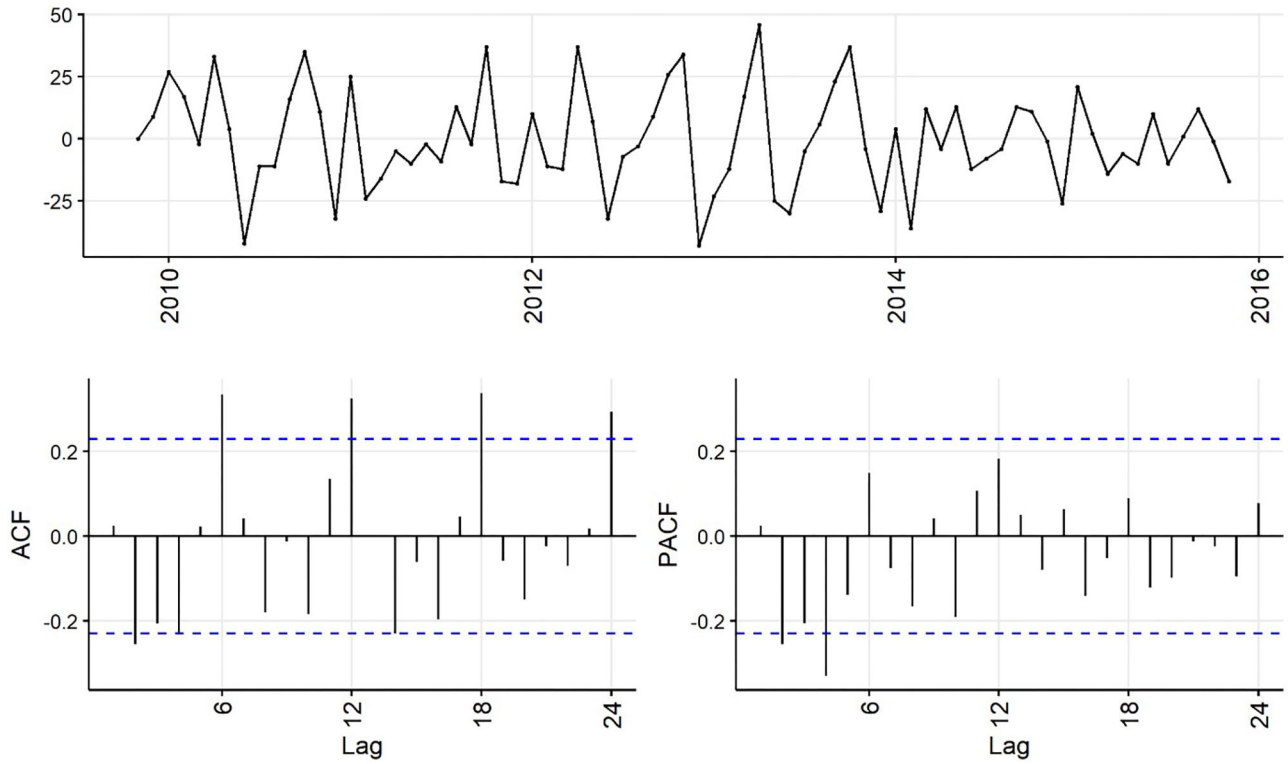


FIGURE 9 | Time-series **(top)**, autocorrelation function **(left)**, and partial autocorrelation function **(right)** plots of differenced monthly aggregated confirmed canine parvovirus events reported in a surveillance system in Australia, 2009–2015.

TABLE 1 | *P*-values of statistical tests for stationarity on weekly and monthly time-series of confirmed canine parvovirus events reported in a surveillance system in Australia, 2009−2015.

	Weekly		Monthly	
Test	**Raw**	**Differenced**	**Raw**	**Differenced**
Ljung-Box	<0.001	<0.001	<0.001	<0.001
ADF	0.86	0.42	0.95	0.60
KPSS trend	>0.1	>0.1	>0.1	>0.1
KPSS level	0.01	>0.1	0.01	>0.1

ADF, augmented Dickey Fuller; KPSS, Kwiatowski-Phillips-Schmidt-Shin. P > 0.05 indicates failure to reject the null hypothesis.

TABLE 2 | Coefficients and 95% confidence intervals for parameters in an ARIMA model fitted to a weekly time-series of confirmed canine parvovirus events reported in a surveillance system in Australia, 2009−2015.

Parameter	**Coefficient**	**95% range**
AR1	0.46	0.41 – 0.52
AR2	0.30	0.23 – 0.37
MA1	−1.00	−1.02 – −0.98
SAR1	0.11	−0.04 – 0.19
Drift	−0.02	−0.04 – −0.00

match the temporal scale of disease causation investigations and management of the disease.

The apparent decrease in reported events during the period—especially toward the end of the series—likely reflects decreasing enthusiasm for reporting in the surveillance system and then an extended period of it being decommissioned. In addition to this long-term trend in the CPV event data, seasonality was apparent. This makes biological sense, since virus survival is affected by climatic factors (25), dog management and behavior can vary with the seasons and human activity, and breeding cycles might add additional seasonality to CPV transmission. Analysis also demonstrates how monthly aggregated data is better than weekly aggregated data (and by implication, daily reported data) for highlighting the seasonal patterns. Removing trend and seasonal components, the remainder of this series had a regularly repeating pattern, indicating that this series of CPV events can be described using an ARMA or ARIMA model. Through a series of documented procedures, ARIMA models fit to the weekly and monthly CPV event series had (generally positive) seasonal and non-seasonal autoregression parameters of order 1 or 2 and a negative non-seasonal moving average of order 1. This indicates that the occurrence of CPV depends on preceding CPV in the relatively short term (prior 1 or 2 weeks or months, or season), modulated negatively

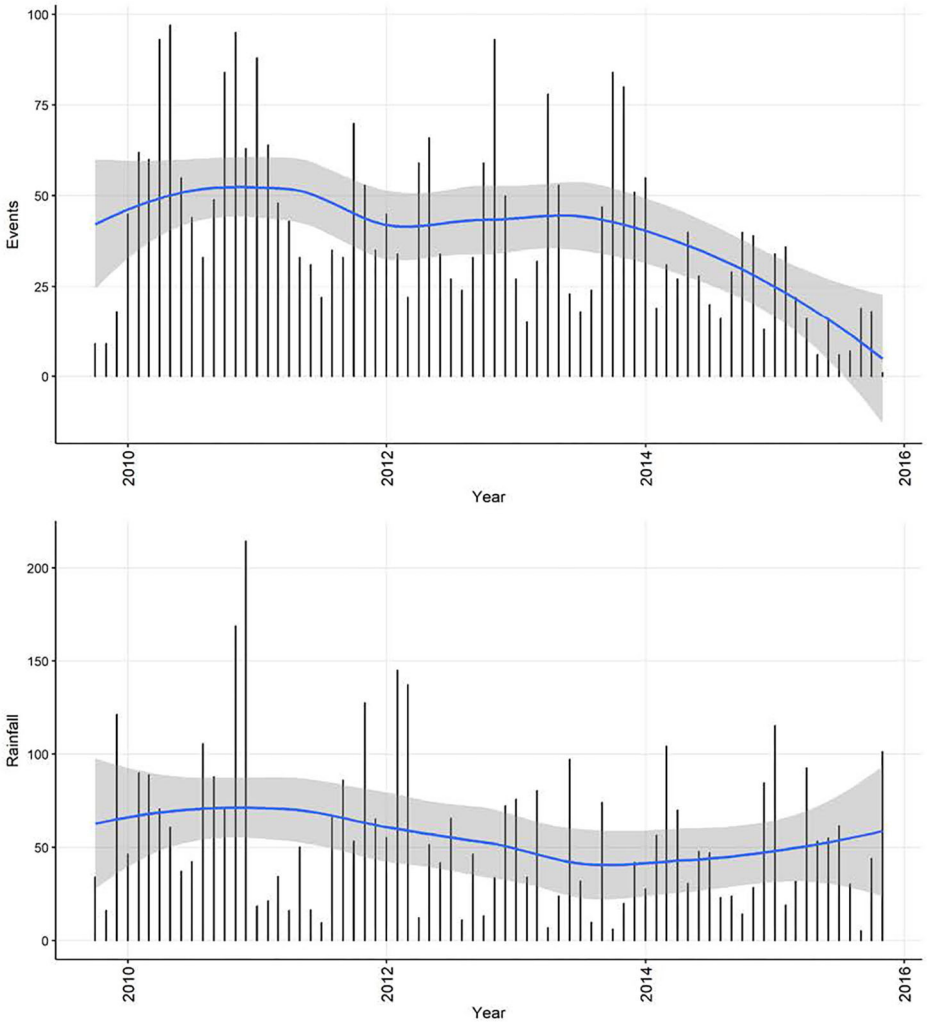


FIGURE 10 | Time-series of confirmed canine parvovirus events reported in a surveillance system in Australia, 2009–2015 **(top)** and corresponding rainfall **(below)** recorded from Mudgee, NSW (32.58°S, 149.58°E). Blue line, loess smoothed curve of events/month with 95% CI (gray).

TABLE 3 | Coefficients and 95% confidence intervals for parameters in an ARIMA model fitted to a monthly time-series of confirmed canine parvovirus events reported in a surveillance system in Australia, 2009−2015.

Parameter	Coefficient	95% range
AR1	0.80	0.51 – 1.08
AR2	−0.29	−0.56 – −0.03
MA1	−0.90	−1.12 – −0.68
SAR1	0.16	−0.09 – 0.39
SAR2	0.29	0.03 – 0.56

by short term variation. This can be interpreted as a disease that responds quickly to recent conditions, consistent with the dynamic transmission of CPV within domestic dog populations. Inclusion of rainfall as a predictor of monthly CPV events did not change the model structure or modify parameter estimates substantially, but indicated that increased rainfall in the previous 2 months and lower rainfall in the month before this is associated with increased number of CPV events reported in the current month. In addition, model fit (AICc) to the data is improved by the inclusion of rainfall, suggesting that rainfall might play a role in the pattern of CPV occurrence. Again, this association might be explained via virus survival or dog behavior (25, 26).

CONCLUSIONS AND RECOMMENDATIONS

To increase the application of methods to analyse time-series data in veterinary epidemiology we recommend that wherever feasible, such time-series data be made available both for analysis and for methods development. We recommend that time-series data be made available, because of those studies identified in our

TABLE 4 | Coefficients and 95% confidence intervals for parameters in a Vector Auto-regressive model fitted to a monthly time-series of confirmed canine parvovirus events and rainfall reported in a surveillance system in Australia, 2009–2015.

Parameter	Coefficient	95% range
AR1	0.58	0.30 – 0.86
AR2	−0.32	−0.56 – −0.08
MA1	−0.79	−0.98 – −0.60
SAR1	0.17	−0.05 – 0.39
SAR2	0.43	0.19 – 0.67
Rainfall 0 m lag	0.07	−0.01 – 0.16
Rainfall 1 m lag	0.12	0.04 – 0.21
Rainfall 2 m lag	0.06	−0.03 – 0.15
Rainfall 3 m lag	−0.10	−0.18 – −0.01

literature scan and reviewed, about one-third described time-series data but failed to use time-series analysis methods; rather, the data were summarized without exploring temporal trends and patterns and autocorrelation. Application of time-series analysis methods has the potential to generate further insight into the occurrence and distribution of animal diseases, disease causation and how it can be used to facilitate surveillance and disease control.

In addition, we recommend that further efforts are made to make analysis of time-series data (whether in R or other software platforms) more user-friendly and accessible. Although lack of availability of data of sufficient length can preclude time-series analysis, lack of familiarity with analytical methods and available software might also limit the information generated by such analyses. In addition, we also recommend that epidemiologic assumptions underlying the analysis of time-series data—particularly a constant population at-risk, non-sparse data, and sources of bias—be thoroughly investigated in veterinary studies. We have not described such investigations here, because they are common to all epidemiologic analyses using observational data.

With developments in monitoring and surveillance systems, and some systems being in existence for extended periods of time, we expect more time-series data to become available together with more software options. However, time-series applications require further promotion to increase adoption and use in veterinary epidemiology. Given that the most often cited advantage of using time-series analyses is the ability to predict disease occurrence, contributing to early warning and therefore disease prevention, application of this analytical method in veterinary epidemiology and preventive medicine is warranted.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article, see (8).

AUTHOR CONTRIBUTIONS

MW conceived the idea and analyzed data from the literature scan. VB developed the R code. MW, RI, and VB reviewed the literature and drafted the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

Dr. Mark Kelman is acknowledged for providing the example data used in this study. We also thank the librarians at The University of Sydney and at Charles Sturt University for sourcing some of the literature.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.00604/full#supplementary-material

REFERENCES

1. Trawinski PR, Mackay DS. Meteorologically conditioned time-series predictions of West Nile virus vector mosquitoes. *Vector Borne Zoonotic Diseases*. (2008) 8:505–22. doi: 10.1089/vbz.2007.0202
2. Ward MP, Carpenter TE. Techniques for analysis of disease clustering in space and in time in veterinary epidemiology. *Preventive Vet Med*. (2000) 45:257–84. doi: 10.1016/S0167-5877(00)00133-1
3. Dorea FC, McEwen BJ, McNab WB, Sanchez J, Revie CW. Syndromic surveillance using veterinary laboratory data: algorithm combination and customization of alerts. *PLoS ONE*. (2013) 8:e82183. doi: 10.1371/journal.pone.0082183
4. Awada L, Tizzani P, Noh SM, Ducrot C, Ntsama F, Caceres P, et al. Global dynamics of highly pathogenic avian influenza outbreaks in poultry between 2005 and 2016 focus on distance and rate of spread. *Transbound Emerg Dis*. (2018) 65:2006–16. doi: 10.1111/tbed.12986
5. Vial F, Wei W, Held L. Methodological challenges to multivariate syndromic surveillance: a case study using Swiss animal health data. *BMC Vet Res*. (2016) 12:288. doi: 10.1186/s12917-016-0914-2
6. Bollig N, Clarke L, Elsmo E, Craven M. Machine learning for syndromic surveillance using veterinary necropsy reports. *PLoS ONE*. (2020) 15:e0228105. doi: 10.1371/journal.pone.0228105
7. Dórea FC, Vial F. Animal health syndromic surveillance: a systematic literature review of the progress in the last 5 years (2011–2016). *Vet Med*. (2016) 7:157–70. doi: 10.2147/VMRR.S90182
8. Brookes VJ, Iglesias R, Ward MP. vbrookes/Timeseries_analysis: ARIMA_TimeSeries. *Zenodo*. (2020). doi: 10.5281/zenodo.3738684. [Epub ahead of print].
9. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna (2019). Available online at: https://www.R-project.org/
10. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag. (2016).
11. Wickham H. The split-apply-combine strategy for data analysis. *J Statistical Software*. (2011) 40:1–29. doi: 10.18637/jss.v040.i01
12. Wickham H, François R, Henry L, Müller K. *dplyr: A Grammar of Data Manipulation*. R package version 0.8.0.1. (2019). Available online at: https://CRAN.R-project.org/package=dplyr
13. Grolemund G, Wickham H. Dates and times made easy with lubridate. *J Statistical Software*. (2011) 40:1–25. doi: 10.18637/jss.v040.i03
14. Trapletti A, Hornik K. *tseries: Time Series Analysis and Computational Finance. R package version* 0.10–46. (2018).
15. Pfaff B. VAR, SVAR and SVEC models: implementation within R Package vars. *J Statistical Software*. (2008) 27:4. doi: 10.18637/jss.v027.i04

16. Pfaff B. *Analysis of Integrated and Cointegrated Time Series with R. Second Edition*. New York, NY: Springer. (2008).
17. Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, et al. *Forecast: Forecasting Functions for Time Series and Linear Models*. R package version 8.7. (2019). Available online at: http://pkg.robjhyndman.com/forecast
18. Hyndman RJ, Khandakar Y. Automatic time series forecasting: the forecast package for R. *J Statistical Software*. (2008) 26:1–22. doi: 10.18637/jss.v027.i03
19. Ward MP, Kelman M. Companion animal disease surveillance: a new solution to an old problem? *Spatial Spatio-temporal Epidemiol*. (2011) 2:147–58. doi: 10.1016/j.sste.2011.07.009
20. Ling M, Norris J, Kelman M, Ward MP. Risk factors for mortality from canine parvoviral-related disease in Australia. *Vet Microbiol*. (2012) 158:280–90. doi: 10.1016/j.vetmic.2012.02.034
21. Brady S, Norris J, Kelman M, Ward MP. Canine parvovirus in Australia: the role of socio-economic factors in disease clusters. *Vet J*. (2012) 193:522–8. doi: 10.1016/j.tvjl.2012.01.025
22. Zourkas E, Kelman M, Ward MP. Canine parvovirus in Australia: a comparative study of reported rural and urban cases. *Vet Microbiol*. (2015) 181:198–203. doi: 10.1016/j.vetmic.2015.10.009
23. Goddard A, Leisewitz AL. Canine parvovirus. *Vet Clin North Am Small Anim Pract*. (2010) 40:1041–53. doi: 10.1016/j.cvsm.2010.07.007
24. Cleveland RB, Cleveland WS, McRae JE, Terpenning I. STL: a seasonal-trend decomposition. *J Off Stat*. (1990) 6:3–73.
25. Rika-Heke T, Kelman M, Ward MP. The relationship between climate, weather and companion animal disease: tick paralysis and canine parvovirus. *Vet J*. (2015) 205:87–92. doi: 10.1016/j.tvjl.2015.03.012
26. Kelman M, Barrs VR, Norris JM, Ward MP. Socioeconomic, geographic, and climatic risk factors for canine parvovirus infection and euthanasia in Australia. *Preventive Vet Med*. (2020) 174:104816. doi: 10.1016/j.prevetmed.2019.104816

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

REVIEW
published: 21 October 2020
doi: 10.3389/fvets.2020.519059

Ecological Niche Modeling: An Introduction for Veterinarians and Epidemiologists

*Luis E. Escobar**

Department of Fish and Wildlife Conservation, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States

Most infectious diseases in animals are not distributed randomly. Instead, diseases in livestock and wildlife are predictable in terms of the geography, time, and species affected. Ecological niche modeling approaches have been crucial to the advancement of our understanding of diversity and diseases distributions. This contribution is an introductory overview to the field of distributional ecology, with emphasis on its application for spatial epidemiology. A new, revised modeling framework is proposed for more detailed and replicable models that account for both the biology of the disease to be modeled and the uncertainty of the data available. Considering that most disease systems need at least two organisms interacting (i.e., host and pathogen), biotic interactions lie at the core of the pathogen's ecological niche. As a result, neglecting interacting organisms in pathogen dynamics (e.g., maintenance, reproduction, and transmission) may limit efforts to forecast disease distributions in veterinary epidemiology. Although limitations of ecological niche modeling are noted, it is clear that the application and value of ecological niche modeling to epidemiology will increase in the future. Potential research lines include the examination of the effects of biotic variables on model performance, assessments of protocols for model calibration in disease systems, and new tools and metrics for robust model evaluation. Epidemiologists aiming to employ ecological niche modeling theory and methods to reconstruct and forecast epidemics should familiarize themselves with ecological literature and must consider multidisciplinary collaborations including veterinarians to develop biologically sound, statistically robust analyses. This review attempts to increase the use of tools from ecology in disease mapping.

Keywords: spatial epidemiology, ecological niche modeling (ENM), disease mapping, ecological niche, distributional ecology

OPEN ACCESS

Edited by:
Victoria J. Brookes, Charles Sturt University, Australia

Reviewed by:
Jaber Amine Belkhiria, University of California, Davis, United States
Jörn Gethmann, Friedrich-Loeffler-Institute, Germany
Gustavo Machado, North Carolina State University, United States

***Correspondence:**
Luis E. Escobar
escobar1@vt.edu

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *10 December 2019*
Accepted: *25 August 2020*
Published: *21 October 2020*

Citation:
Escobar LE (2020) Ecological Niche Modeling: An Introduction for Veterinarians and Epidemiologists. Front. Vet. Sci. 7:519059. doi: 10.3389/fvets.2020.519059

INTRODUCTION

Spatial epidemiology is the branch of epidemiology that aims to understand the geographic distribution of diseases (including its causative agents, hosts, and related factors) (1, 2). Most diseases in animals are not distributed randomly across landscapes or regions. Instead, researchers can quantitatively determine specific environmental factors associated with the occurrence of disease (3, 4). Reports of the spatial location of pathogens, disease vectors, or reservoirs are becoming more abundant, high quality, and openly accessible for a series of infectious diseases. Similarly, data on environmental variables are increasing in availability and cover diverse spatial and temporal scales: from meters to continents and from days to centuries (both retrospective

and predictive). For example, many datasets of soil composition and structure (5), landscape composition and structure (6), and climate and geomorphology (7–9) are freely and openly available for mapping diseases in aquatic and terrestrial ecosystems globally. These variables can be linked with disease data to reconstruct or predict the geographic distribution of environmentally (e.g., anthrax), vector-borne (e.g., Bluetongue disease), and directly transmitted diseases (e.g., rabies), which are important to veterinary medicine. These studies, however, need a basic understanding of Geographic Information Systems, spatial statistics, and a deep understanding of the biology of the disease system to be modeled.

Disease models accounting for environmental information are particularly informative to understand the spread of diseases that are undergoing range expansion, which is termed "distributional disequilibrium" in ecology (10–12). Ecological theories and methods are commonly used in spatial epidemiology to design and interpret models of conditions where infections are likely to occur, with outputs projected to geography as measures of "suitability." Suitability has been defined as the "sum" of the effects of resource and environmental conditions on the fecundity, demography, and survivorship of populations (13). Ecological niche modeling has been the main branch of ecology employed to map disease transmission. Comprehensive reviews are available elsewhere regarding the fundamentals of ecological niche modeling for epidemiologists interested in its applications on medical geography of infectious diseases (14–16). This manuscript is an overview of the field of ecological niche modeling for veterinarians and epidemiologists and considers parasites (e.g., tapeworm) and pathogens (e.g., virus) as agents causing disease. The content of this review is a friendly introduction to more specialized literature and study cases described in more detail elsewhere (17, 18).

Models

A model is a simplification of a complex system. For example, in biomedicine, mice could be used as animal models to understand the effects of a drug in humans. In mathematics and statistics, equations can be used to simplify and summarize complex phenomena. Some mathematical models can be complex, by accounting for many details (i.e., parameters) in the disease system, while other models can be simple, accounting for just a few, key components of the system. Models can be used to reconstruct the structure or functioning of the system in question—termed descriptive models (e.g., the specific temperatures where a disease vector is found). Complementarily, models could be used to anticipate how the system would respond to determined "what-if-scenarios" —termed predictive models (e.g., the expected distribution of a disease vector under future temperature). Descriptive models are the basis and first stage for the development of predictive models.

Descriptive models are generally evaluated in terms of the capacity of the model to accurately reconstruct patterns found in the available data. Thus, evaluation metrics used to differentiate between good and bad descriptive models generally account for the amount of information lost (e.g., Akaike's information criterion) (19) (**Figure 1A**). Predictive models are evaluated based on their capacities to accurately predict, better than by random, new data (i.e., independent data not used during model calibration). Therefore, evaluation metrics used to differentiate between good and bad predictive models commonly measure model capacity to differentiation between actual data and random observations (e.g., *p*-value, sensitivity vs. specificity) (**Figure 1B**).

Models can also be differentiated based on their capacities of interpolation and extrapolation. Interpolation is defined as an estimation of unknown values present within the range of values from the data used to calibrate the model (20). Extrapolation is defined as the estimation of unknown values beyond the range of data used for calibration. Ideally, models aiming to be descriptive should have low interpolation and extrapolation abilities resembling good fit to the data. Predictive models are expected to interpolate and extrapolate. As a result, the final goal of the model, descriptive or predictive, should guide the design of its calibration and evaluation protocols. A perilous arena within spatial epidemiology is the development of predictive models that are evaluated using metrics developed for descriptive models or that are penalized based on extrapolation (21). Similarly, robust model evaluation of predictive models would require evaluation data statistically independent from calibration data. Thus, data-patitioning methods that do not ensure independency (e.g., cross-validation) have questionable capacity to differentiate between good and bad predictive models.

ECOLOGICAL NICHE MODELS

Previous applications of ecology to map infectious disease risk have resulted in successful disease control and prevention [e.g., (22)]. During the last two decades, valuable advancements in alternative approaches to investigate infectious diseases through applied ecology have been made (15). The ability to determine why a disease is present in one animal species, season, and geographic area but absent in others facilitates the understanding of spread and persistence of infectious diseases in wildlife and domestic animal populations, critical for veterinary medicine.

The final goal of ecological niche modeling applications in spatial epidemiology is to determine environmental conditions associated with disease occurrence. This in turn can help to identify localities where such conditions exist and that are suitable for disease introduction, maintenance, and posterior spread. These models can be conducted at the local level using accurate disease reports coupled with landscape information or at the regional level coupled with climatic variables.

Disease distributions at coarse scales are often manifested through climatic variables (e.g., temperature and precipitation) falling across expected ranges of climate values observed in the bulk of confirmed disease reports. In ecological niche theory, the fundamental niche, $\mathbf{N}_F$, represents the set of abiotic environmental conditions necessary for long-term population persistence. More specifically, $\mathbf{N}_F$ allows population permanence without subsidy from immigration. Variables used to estimate $\mathbf{N}_F$ are not modified by the presence or abundance of the organism (e.g., temperature, precipitation) (14). $\mathbf{N}_F$ models are usually estimated at coarse-scale based on climatic signatures of biological systems to

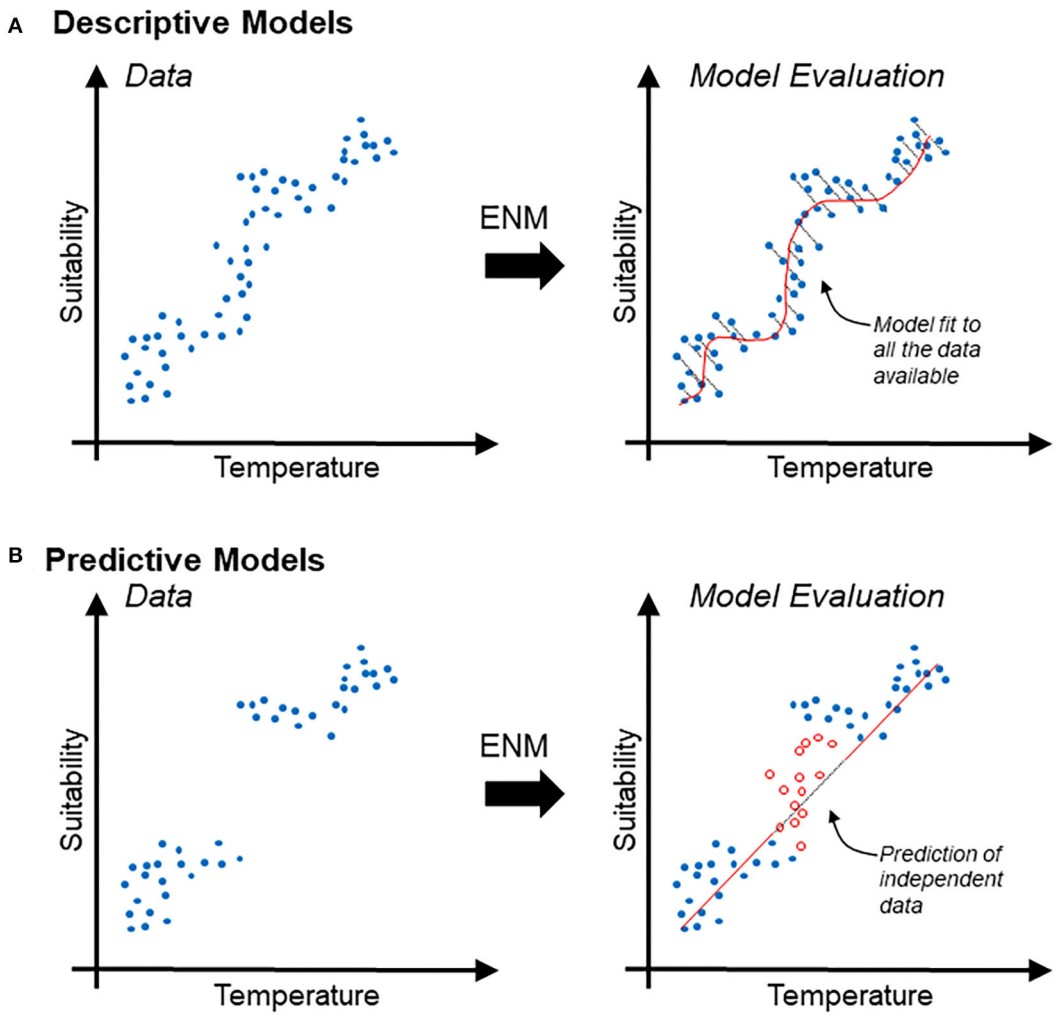


FIGURE 1 | Schematic description of descriptive vs. predictive models in the context of ecological niche modeling. **(A)** Disease occurrence data (blue points) used to estimate relationships between temperature (independent variable; *x*-axis) and suitability (dependent variable; *y*-axis), interpreted as the probability of specific environmental combinations to mirror the conditions where the species actually occurs. Correlative ecological niche model (ENM) can also be based on logistic regression based on relation between continuous environmental variables and binary reports of the species (i.e., presence/absence). Note that the descriptive model (red line) is evaluated in terms of its capacity to accurately resemble the data; the information lost is expressed as the distance between the model and the data employed for model calibration (right). **(B)** Predictive model (same as above) intended to forecast the response of the system to an unknown status. Predictive models are generally evaluated based on their capacity to predict independent data (i.e., data not used during model calibration). Note that a predictive model could be simple (red straight line) and could result in the loss of more information. Nevertheless, independent data (red points) may be accurately predicted.

reconstruct the potential geographic distribution of organisms, revealing areas with suitable climatic conditions across broad regions (14).

Empirical and theoretical evidence from physiological experiments suggests that population growth and survival of species often have a Gaussian response to environmental gradients (23–28). That is, theory suggests that an organism's fitness responds to environmental conditions with a normal curve, where extremely low and extremely high environmental values drive low fitness, while intermediate environmental values are the optimum for fitness (**Figure 2A**). $\mathbf{N}_F$ accounts for multiple environmental variables, and when many environmental variables are considered, each with a Gaussian response curve, their combination could resemble an ellipsoid (**Figure 2B**). Consequently, Maguire proposed that the $\mathbf{N}_F$ should be convex in shape (25), with ellipsoids offering simple proxies of such convex shape (29). The Gaussian response of organisms to environmental conditions suggests that a disease reservoir or vector would have varied demographic parameters in different sections of its $\mathbf{N}_F$. As a result, disease control on reservoirs or vectors (e.g., culling, vaccination) could have different effects under alternative environmental conditions. More specifically, this theory suggests that highest transmission should be expected in the optimal environmental conditions suitable for a disease reservoir or vector (30) and population health management decisions should be made accordingly.

Nevertheless, all the environmental conditions in $\mathbf{N}_F$ may not be entirely available for the species. Thus, $\mathbf{N}_F$ is hard to

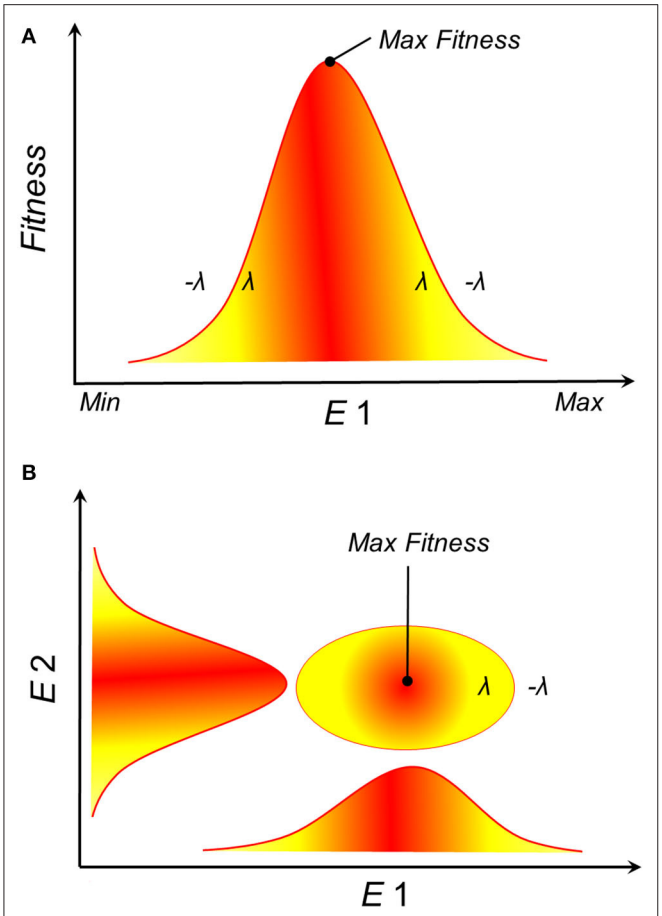


FIGURE 2 | Theoretical representation of the fundamental ecological niche. **(A)** Survival or population growth (*y*-axis), as a proxy of suitability, shows a Gaussian response to environmental gradients (*E*1, *x*-axis). For example, population growth (**λ**) is positive within the environmental range tolerated by the species and negative below or above such range (–**λ**), with an optimum observed at intermediate environmental values (red). **(B)** Ellipsoidal shapes are the expected species response to *n*-dimensional environmental gradients. Example in a bidimensional environmental space where red is higher suitability.

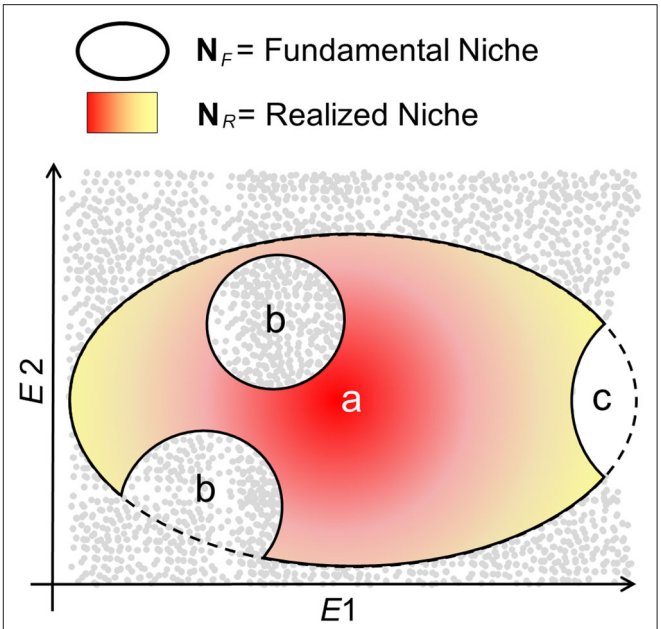


FIGURE 3 | Schematic of N_R vs. N_F in a disease system. The fundamental niche, $\mathbf{N}_F$, denotes the environmental conditions, available (gray points) or not (white), that are suitable for the organism (e.g., pathogen, vector, and reservoir) to persist and establish populations in the long term. $\mathbf{N}_F$ is represented as an ellipsoid with *a* denoting the theoretically most suitable conditions (dark red), with suitability declining to the edges of $\mathbf{N}_F$ (yellow). The realized niche, $\mathbf{N}_R$, denotes the environments in $\mathbf{N}_F$ actually occupied by the organism. Nevertheless, $\mathbf{N}_R$ may be a subset of $\mathbf{N}_F$ maybe because the organism does not occupy some environments due to dispersal limitations (*b*), because there is not host of vector for the organism's persistence, or because the suitable environmental conditions do not exist in the area accessible (*c*), reducing $\mathbf{N}_F$ to $\mathbf{N}_F{}^{\star}$.

reconstruct with field data due to its theoretical nature; however, other portions of it may be more feasible to estimate. The realized niche, $\mathbf{N}_R$, is the portion of the $\mathbf{N}_F$ that is actually occupied by the organism, reflecting fine-scale constraining effects of dispersal limitations and biotic interactions (14). Fine-scale ecological niche modeling is generally achieved by linking landscape-level variables (e.g., vegetation, host density) as a proxy of the $\mathbf{N}_R$. Essentially, $\mathbf{N}_R$ is a close representation of the actual conditions present across the distribution of an organism. Recent experimental research shows that landscape materials (e.g., grass, wood, soil, and water) can play a role in the maintenance and spread of pathogens facilitating environmental transmission (31, 32). Satellite-derived data of vegetation phenology, soil composition, moisture, and microclimate have served as proxies of landscape features, allowing researchers to capture environmental signatures of pathogen distribution at the local level (6, 33).

Ecological niche models assume that the biotic interactions restrict species to occupy their entire $\mathbf{N}_F$; therefore, that $\mathbf{N}_R$ is a portion of $\mathbf{N}_F$ (14, 34) (**Figure 3**). That is, maybe not all the utilizable conditions in the pathogen's $\mathbf{N}_F$ have the presence of the host. Additionally, ecological niche inferences must consider the geographic area accessible to the organism (termed **M**) and the set of environments represented across that region (14) as the limits of $\mathbf{N}_F$ that are available. That is, maybe not all the climates utilizable by the organism exist in the areas of study so that the existing fundamental niche, $\mathbf{N}_F^*$, represents the portion of $\mathbf{N}_F$ that the species could use. $\mathbf{N}_F^*$ is the intersection of $\mathbf{N}_F$ with the area accessible **M**, such that the existing $\mathbf{N}_F^*$ will be a subset of $\mathbf{N}_F$; any attempt to use the existing $\mathbf{N}_F^*$ or $\mathbf{N}_R$ as estimates of $\mathbf{N}_F$ is perilous for species with limited dispersal (35). Based on this reasoning, the selection of the study area has dramatic implications on the environmental conditions to be modeled and, in turn, on estimations of $\mathbf{N}_F$ or $\mathbf{N}_R$.

THE PROBLEM OF SCALE IN INFECTIOUS DISEASE ECOLOGY

A major challenge in spatial epidemiology and distributional ecology is the identification of the scale of the analysis for

a statistically correct study design and a biologically sound model interpretation. While this question may appear to be easily answered based on the data available, generally, incorrect identification of the scale may result in misleading study designs, misinterpretation of results, and inability to fill primary gaps of knowledge. Studies must identify the temporal scale (from hours and days to decades and millennia, including past, present, and future time) and the spatial scale (from centimeters to kilometers) of interest. For example, studies could be conducted in a protected area during a season, or at the continental level across a 60-years period. It is also important to define the organismal level (from molecules and genes to populations and biomes) upon which the study is focused (36). Because infectious diseases can be examined on a wide scale, from micro to macro, assumptions, data, and model interpretation will vary across scales (**Figure 4**).

A main assumption in epidemiology is that diseases do not occur randomly, which can be used to assess the distribution of pathogens across taxa and geographies to identify specific patterns that can be modeled and predicted. At the fine scale, models can estimate the likelihood that specific wildlife species will be suitable for vector infestation. At the medium scale, models can assess the landscape drivers of disease transmission. At the coarse scale, models could be used to reconstruct spatial patterns of the extent, direction, and speed of disease spread across continents (**Figure 4**).

Fine-scale studies are conducted locally and capture individual-level details in short periods of time (e.g., a season). The resources and effort necessary to conduct fine-scale research restrict their development to small study areas (e.g., a forest). Coarse-scale studies, however, can be conducted at large extents but generally fail to capture the details necessary to understand local-level phenomena. The level of detail or grain of variables quantified is linked to the scale, extent, and their capacity of prediction. Coarse-scale studies may lack details but would provide predictions that are more robust across space and time. Thus, the problem of scale in disease ecology is how predictions change as scales change (**Figure 5**). The problem of scale (i.e., temporal or spatial) has been described in detail by Simon A. Levin (36) and provides opportunities to better understand disease systems across space and time. Interestingly, spatial epidemiology of animal diseases seems to be biased toward local-level studies, with limited research conducted at coarser scales (37).

The organismal level is a major challenge in spatial epidemiology. For instance, the complex transmission cycles of vector-borne, water-borne, or directly transmitted diseases require two (e.g., pathogen and host) or more (e.g., pathogen, multiple vectors, multiple reservoirs, or hosts) species to be included in the model. Often in practice, modelers use a single organism to reconstruct areas of transmission, which could focus on the vector or the pathogen. While a parsimonious approach, it requires a strong understanding of the ecology of the disease in question to identify the organism that best explains the disease system. Thus, a next frontier in ecological niche modeling applications to disease systems is the inclusion of more biological components of the cycle of transmission in the modeling process.

ECOLOGICAL NICHE MODELING AND SPATIAL EPIDEMIOLOGY

Ecological niche modeling has proven to be a useful tool for forecasting distributions and distributional changes for a vast number of organisms (22, 38, 39) and is increasingly employed to predict distributions of pathogens on diverse spatial scales (15). Traditional ecological niche modeling frameworks, however, may make unrealistic assumptions and therefore yield inaccurate predictions. These modeling frameworks must therefore be revised and amended if they are to work in epidemiology (40). Ecological niche modeling estimates ecological niches of species by linking spatial occurrence records with environmental covariates, *via* correlative or mechanistic approaches (41, 42). Theory and analytical approaches of ecological niches have been described during the last century (43), especially for biodiversity and conservation studies.

A decade ago, it was hypothesized that coarse-scale geographic distributions of species were constrained principally by abiotic environmental conditions (i.e., inert variables) across relevant regions, with biotic interactions having negligible effect [termed the *Eltonian Noise Hypothesis* (14, 29)]. As a result, most modelers have considered it reasonable to assume that influences of biotic interactions could be neglected in ecological niche modeling (29). The Eltonian Noise Hypothesis, however, was conceived in the context of free-living organisms (e.g., plants, birds). Currently, most ecological niche modeling applications do not include biotic variables (i.e., derived from living organisms). Epidemiologists of infectious disease and veterinarians have a clear understanding of the major flaws of models neglecting biotic interactions because infectious diseases are by definition biotic interactions.

Developing models based solely on abiotic variables make model outputs of easy interpretation. For example, a model based on pH and humidity could generate estimates of suitability with regard to environmental conditions. Nevertheless, the role of biotic variables has not been assessed rigorously in parallel analyses in disease ecology [but see (44)]. Indeed, incorporation of biotic variables in ecological niche modeling analyses for diseases was proposed only relatively recently (45), and such applications remain rare in spatial epidemiology. The inclusion criteria of the biotic variables to be used, their temporal and spatial scales, and whether biotic variables should be used before, during, or after the model calibration process remain understudied (40). Currently, use of abiotic-only predictors (e.g., climate) dominates the literature regarding modeling and predicting geographic distributions of pathogens. Including biotic variables in the ecological niche modeling process would require a detailed and *a priori* definition of the modeling outputs. For example, a model including host density or percentage of vaccination coverage would require a revision of the "suitability" term in the context of each study (e.g., suitability for transmission or exposure); alternatively, other terms would need to be employed for modeling disease systems, such as risk (46) or relative occurrence rate (47). Understanding the role of biotic

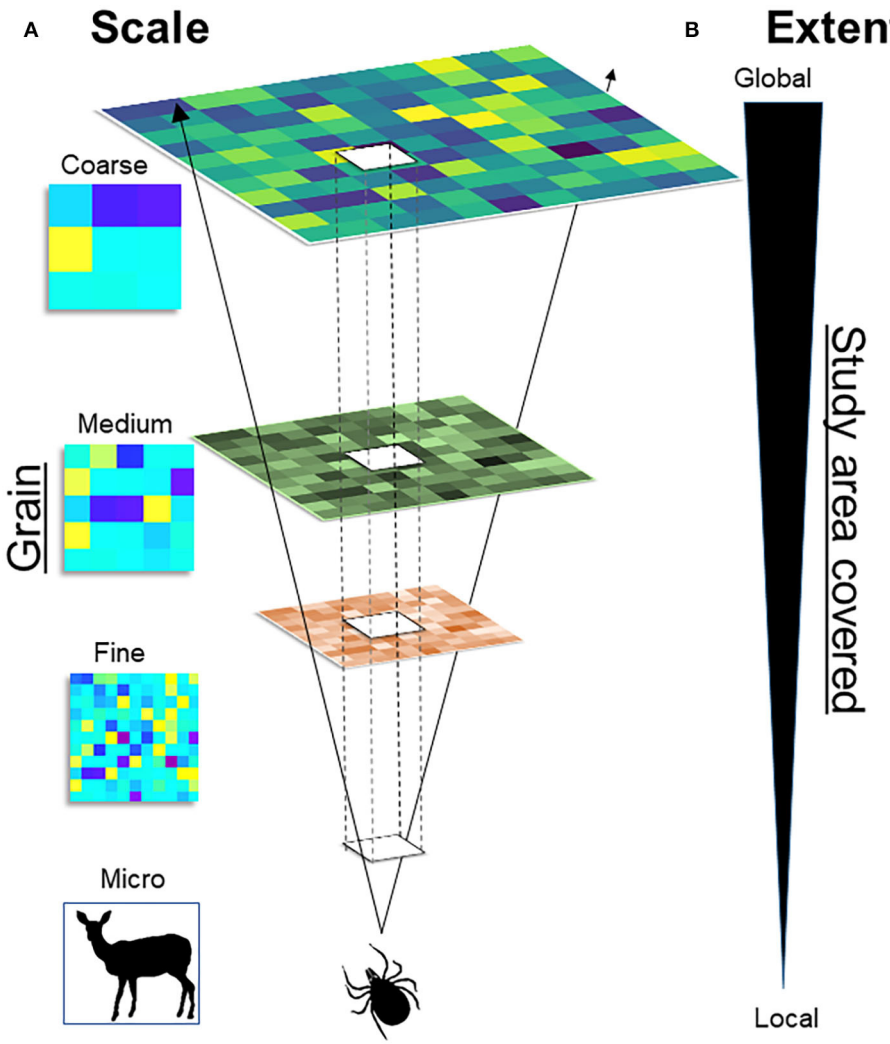


FIGURE 4 | Multiscale framework. **(A)** Scale: Scale of variables (i.e., temporal and spatial) varies from the micro, to the fine, to the coarse. For example, ticks could be studied at the micro scale, by assessing its distribution across the skin of the host (e.g., deer). **(B)** Extent: Represents the size of the study area, from the local to the global extent. For example, increasing the extent will allow to study ticks across a forest (local) or across a continent (global). The scale and the extent are correlated: Fine-scale studies provide high detail (fine grain) but cover small study areas, coarse-scale studies cover large areas at the cost of detail (large grain).

variables in ecological niche modeling may revolutionize the utility of these tools drawn from ecology for disease risk mapping.

INGREDIENTS TO MAKE AN ECOLOGICAL NICHE MODEL

Historically, "ingredients" to build ecological niche models have been summarized in three major categories: occurrence data, environmental data (abiotic or biotic variables), and algorithm (**Figure 6**). Occurrence data are represented as disease cases, or serology or direct detection of pathogens or parasites, or records of vectors, intermediate hosts, or reservoirs recorded geographically as coordinates (i.e., latitude and longitude). Environmental data are represented at coarse (i.e., climate) and fine (e.g., vegetation indices) resolutions in terms of the abiotic and biotic environmental conditions where occurrence data are collected. Then, to link environmental conditions and disease occurrence, correlative or classification algorithms are generally used. This analytical framework has been criticized due to the limited understanding of the user regarding the potential data and algorithm limitations, and the theoretical bases of the algorithm employed (48). Careless applications of this simple modeling framework has been termed "click-and-run ecological niche modeling" (49) and has resulted in misleading ecological niche modeling applications (50–52). Indeed, studies to reconstruct disease distributions should avoid using protocols and parameterization scenarios developed for other taxa, regions, or periods. Instead, the modeling protocol for disease mapping should be specific to the study question, data available, and assumptions of the disease system.

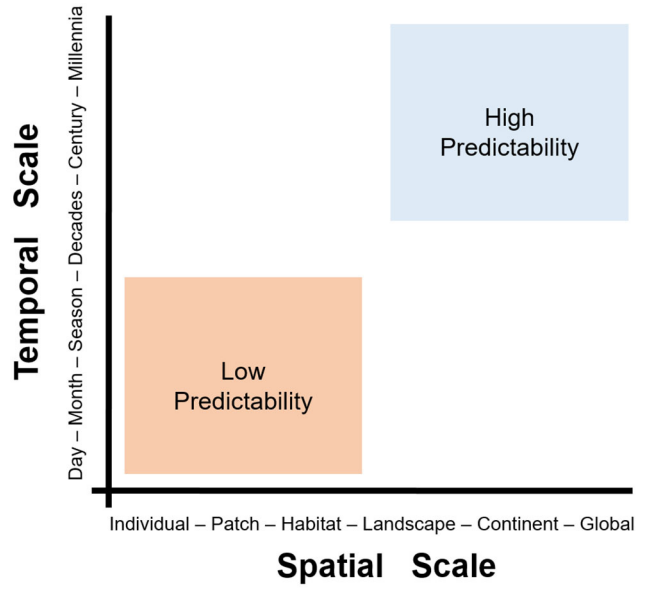


FIGURE 5 | The problem of scale in disease ecology. Scale can be defined in two dimensions: spatial (*x*-axis) and temporal (*y*-axis). Fine-scale studies (e.g., data from at the individual-level during hours to days) can provide high-quality data and a good understanding of phenomena. The figure shows a pragmatic categorization of two spatiotemporal scales, fine (orange) and coarse (blue). Fine-scale studies (orange) tend to offer limited predictability but high detail, while coarse-scale studies (blue) offer high predictive potential at the cost of losing detail. Note that landscape/seasonal studies are somewhere in-between fine- and coarse-scale studies. Modified from Levin (36).

The click-and-run modeling framework (**Figure 6**) is based on the use of "recipes" to model the distribution of any species, neglecting the biology of the organism in question. These models also neglect biases or artifacts in the data used for model calibration and the functionality of the algorithm employed. This approach requires limited data curation and model parameterization and was used in the past for single-species ecological niche models, but is currently used for studies modeling hundreds or thousands of species to capture coarse ecological patterns (i.e., macroecology). Model evaluation in click-and-run modeling is generally poor or absent, making this modeling framework particularly questionable when modeling infectious diseases.

Models should include occurrence data curated carefully to include only trusted occurrence records for model calibration. Trustworthy disease occurrence records should have traceable diagnostic methods, data sources, transparent surveillance protocols, temporal details, and quantified uncertainty (e.g., spatially error, sensitivity of the diagnostic method). That is, selection of the occurrence data should include an exhaustive inspection of the metadata to reduce errors followed by estimations and mitigations of duplicates, autocorrelation, and sampling bias, supported by detailed protocols as described by Cobos et al. (53).

Ecological niche modeling of disease systems should consider abiotic environmental variables (e.g., temperature, soil, precipitation) that fit the scale of interest and the biology of the disease to be modeled. For example, historical satellite-derived bioclimatic data can be employed at a pixel resolution of 20 km, if this is in agreement with the approximate home range size of a pathogen's reservoir (∼10 km) (10). In this case, one could use land surface temperature (°C) and ground humidity (kg) to overcome limitations of interpolated climatic data (8). That is, it has been found that satellite-derived data overcome limitations of the original interpolations found in the climatic data from ground stations (54). For example, WorldClim, a commonly used resource for climate data, often includes only 0.001% of empirical data and 99.99% of interpolated data, resulting in high spatial lag (autocorrelation) and frequent aberrant and unrealistic climatic values (15).

Biotic variables (e.g., host density, prey density, and predator occupancy) are biological factors shaping the distribution of a species or disease at the local level (55, 56). Biotic variables must be selected on the basis of the natural history of the target pathogen, including the presence or abundance of other organisms that facilitate or limit its presence. For some pathogens, relevant positive biotic factors that facilitate the presence of disease may include co-infections with other pathogens, vegetation preferences of vectors, and host availability, behavior, and density. Negative biotic factors that limit pathogen circulation and establishment may include host immunity and biodiversity values (40, 46, 57). Biotic components may be critical to understand the ecology of pathogen transmission, and their effects are evident when developing studies at fine geographic scales (46), although the question of their action across broad geographic extents remains unanswered. Nevertheless, based on the biology of the pathogen, biotic variables could include proxies of host availability (58), anthropogenic disturbance (59), wildlife reservoirs availability (60), and barriers of disease spread (61, 62). Each of these dimensions has been found to be predictors of infectious diseases (10, 45, 63–67).

The inclusion of biotic variables in ecological niche modeling could be done before, during, and after the calibration of the model. For example, biotic variables could be used before the development of the model by restricting the distribution of the focal species (e.g., pathogen) to regions where biotic interactions may occur (e.g., host distribution; pre-processing). Biotic variables could be added to model calibration by incorporating biotic factors as predictor variables in the ecological niche modeling (e.g., host density; processing). Alternatively, biotic variables could be used once the model is developed by incorporating biotic variables on the final model output (post-processing). For example, a hypothetical model to estimate transmission risk of rabies (*Lyssavirus*) transmitted by vampire bats (*Desmodus rotundus*) at the local level could include the use of abiotic (e.g., temperature and precipitation) and biotic variables. Biotic variables could include livestock densities as proxy for food resources for the vampire bats (58), surface of roads as proxy of local-scale barriers (61, 62), and satellite-derived nighttime light surface as proxy of populated centers (59), since these variables have been proposed as predictors of rabies in wildlife (10, 63–67). That is, when biotic variables are included to reconstruct a disease system, it is crucial to identify

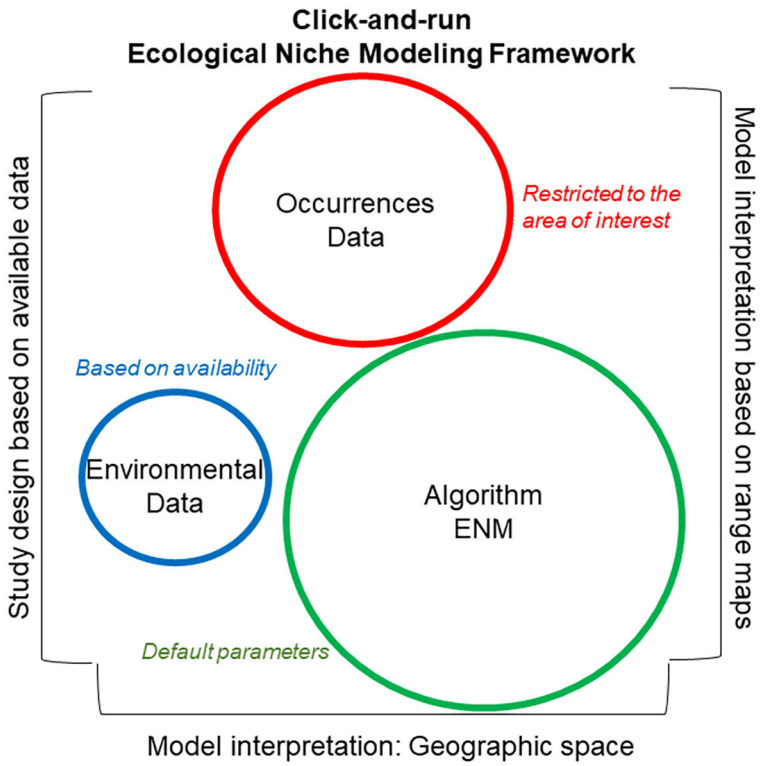


FIGURE 6 | Click-and-run ecological niche modeling framework. This modeling framework is data-driven, with uncaring study designs. Syndromes used in identifying click-and-run ecological niche models include (i) strong emphasis on the good performance of the algorithm employed based on external literature and use of default parameters (green), (ii) moderate emphasis on the quality and quantity of the disease localities (i.e., occurrence data; red), with limited data curation and evaluation, and (iii) little or no justification and defense of the environmental variables selected for model calibration (blue). Additionally, these models generally fail to define the ecological niche to be estimated, $\mathbf{N}_F$ vs. $\mathbf{N}_R$. The study area of these models is generally delimited based on data availability or administrative boundaries, while model interpretation is generally based on visual inspection of maps. Circle size denotes magnitude of importance assigned to each category.

key factors that directly facilitate or limit transmission. Using biotic variables in ecological niche modeling is still not a common practice and more research is necessary in this area to develop a revised modeling framework.

A revised ecological niche modeling framework could facilitate replicable estimations for any disease system (**Figure 7**). Nevertheless, each component of a revised modeling framework (i.e., occurrences, environmental variables, and modeling algorithm) would require careful inspection to discard noise signals due to incorrect study designs. That is, study designs should be based on biologically justifiable study areas and variables, which are important drivers of ecological niche modeling performance. In some situations, the protocol will allow one to determine if a robust ecological niche model is feasible or not.

An important component of the revised protocol is the careful inspection of occurrences using a specified inclusion criterion that prioritizes quality over quantity. To assess and mitigate sampling bias in disease reports, modelers can use the method proposed by Varela et al. (68), which compares models from different occurrences filtering methods to mitigate both oversampled areas and oversampled environments. This approach allows the generation of a series of models under different bias mitigation scenarios to (i) reduce model overfitting (i.e., models mirroring closely the data, resulting on limited learning from the model compared with the raw data) and (ii) capture variability for more informed model interpretations. This methodology has been employed broadly to study the distribution of biodiversity but has been barely used to model infectious diseases. In the revised protocol (**Figure 7**), model calibration could include biotic variables as predictor (45, 69). Nevertheless, researchers must clarify the units and interpretation of the modeling output.

Finally, the study area of interest [**M** *sensu* (70)] is a major component of the modeling process. A common failure in ecological niche modeling applications based on correlative models is to pragmatically determine the study area. Restricting models based on administrative areas (e.g., municipality, department, province, and state) does not account for the biology of the organism. Pathogens do not know about political borders; therefore, models should account for biogeographic barriers (e.g., rivers, roads, impervious surfaces, and oceans) for biologically sound study designs.

The perils of careless study-area delimitations will result in models that are misaligned with the primary question of the study, the ecology of the organism, resulting in underestimations of the true potential of the disease spread. For example, the mosquitoes *Aedes aegypti* and *Ae. albopictus*

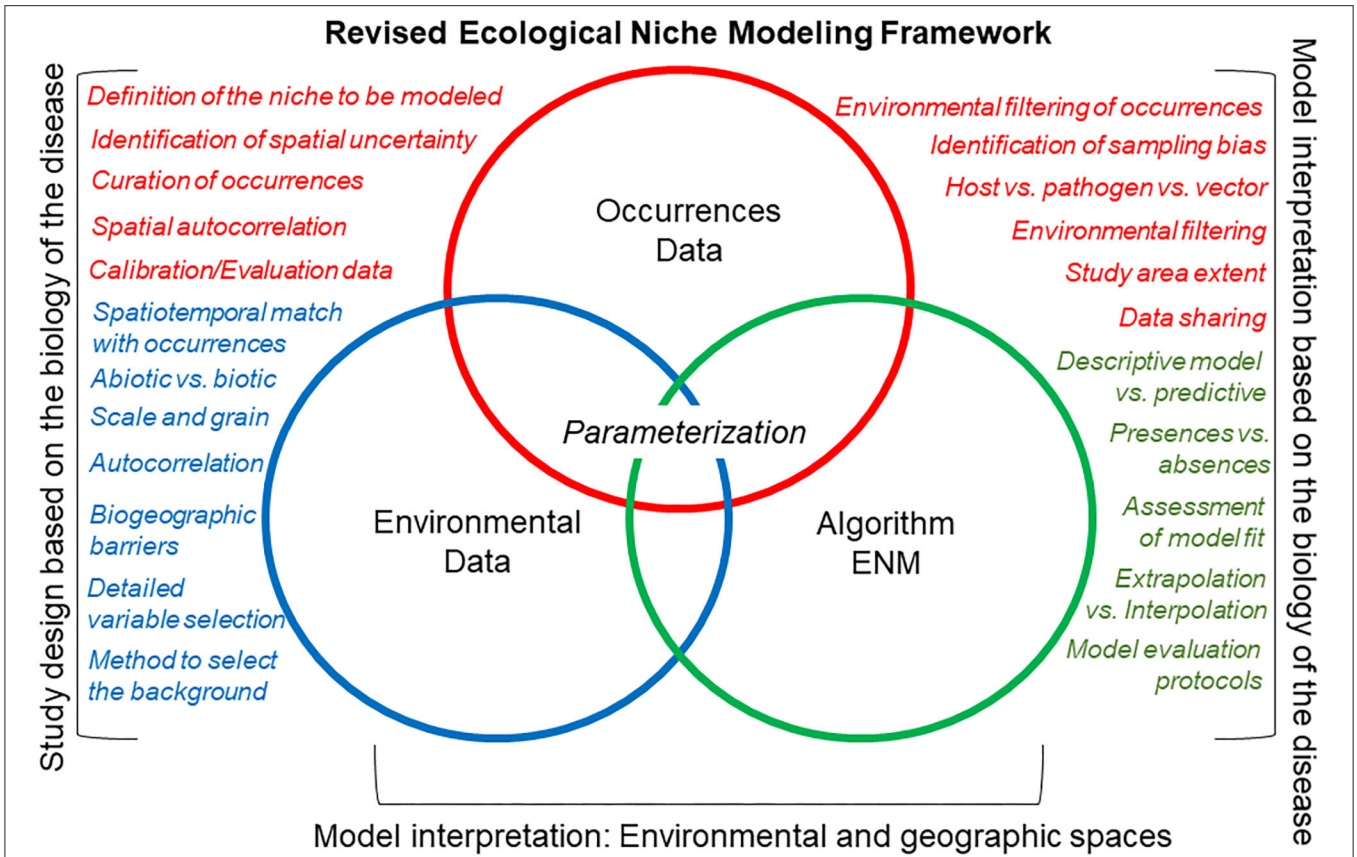


FIGURE 7 | Canonical ecological niche modeling framework (a.k.a. "Peterson's Framework"). This revised ecological niche modeling framework includes some of the critical decisions necessary to develop a comprehensive, replicable ecological niche model that accounts for the natural history of the organism, the quality of the data, and the goal of the model. Here, occurrence data (red), environmental data (green), and the integrative algorithm (blue) have comparable importance regarding their assessment during the modeling process (i.e., same size of circles). The main differences of this canonical ecological niche modeling framework, as compared with the classic ecological niche modeling framework (**Figure 6**), are that more details are evaluated for each modeling component (circles), model design, and interpretation account for the ecology of the organism and data available, and that model outputs are interpreted in geographic and environmental dimensions.

are important vectors of many arboviruses, including Zika, Dengue, and Chikungunya viruses with transmission reported globally. Nevertheless, one may be interested in modeling the distributional ecology of these mosquitoes in a specific study area (**Figure 8A**). For example, mosquito presence in Guatemala along with climate information from the sampling region will only capture environmental tolerances of the species in that particular area. This may therefore result in a gross underestimation of the true potential distribution of the vectors and the diseases they transmit. Indeed, these mosquito species are ecological generalist species that tolerate a broad range of climatic conditions and have global distributions (**Figures 8B,C**). Thus, real tolerances and actual potential distribution of species could be masked by a restricted study area that only accounts for a portion of the species truly potential.

ECOLOGICAL NICHE MODELING ALGORITHMS AND TOOLS

Modeling algorithms in ecological niche modeling have been described elsewhere (47, 72–74), generating starting points for new modelers. Algorithms to develop ecological niche models can be divided into three categories: presence-absence, presence-background, and presence-only. Presence-absence algorithms need a set of localities where the organism occurs (i.e., presence) and a set of localities where the organisms does not occur (i.e., absence). Presence-absence models are calibrated by comparing environmental conditions where the organism is present vs. where it is absent and are generally useful to reconstruct the distribution of diseases at fine scale and short periods, resulting in the need of accurate localities and high-resolution environmental variables. These models, however, have limited capacities to be projected to different areas or periods, instead, their signals are space and time specific. Many algorithms are available including regression (e.g., Generalized Linear Models and Generalized Additive Models) (**Figure 1**) and classification (e.g., Boosted Regression Trees, Random Forest, and Support Vector Machines) (**Figure 9A**) algorithms, with protocols described in detail elsewhere (75).

Occurrence data are generally robust, while absence data are largely questionable in quality and of limited availability [discussed in (14)]. To solve this problem, researchers generally

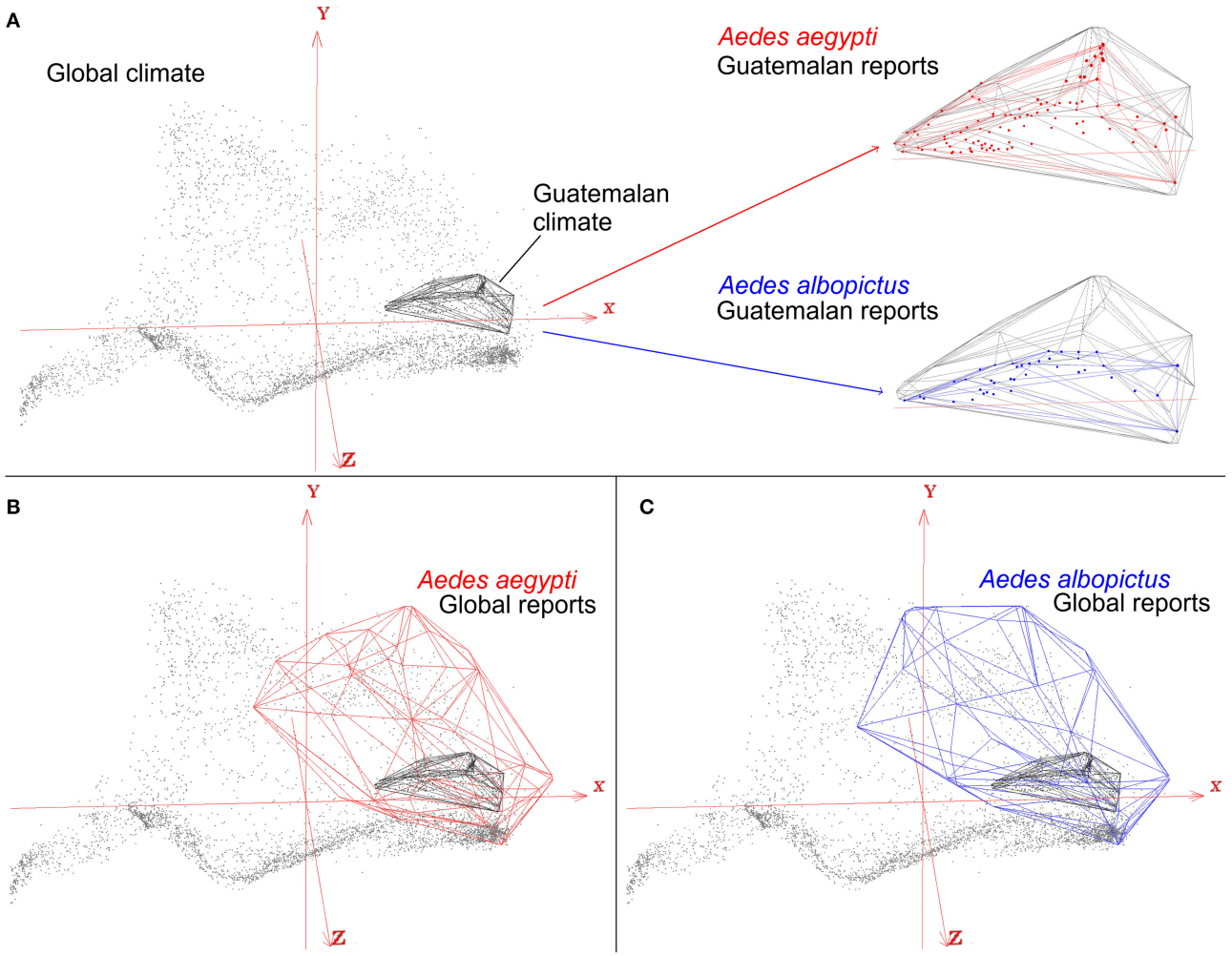


FIGURE 8 | Effects of study area M on ecological niche modeling results. An example of an ecological niche model of arbovirus vectors *Aedes aegypti* (red) and *Ae. albopictus* (blue), which occur worldwide, to predict transmission risk in Guatemala. Global climate data (gray dots) show the broad range and diversity of global environmental conditions summarized in the first three components of bioclimatic variables (red axis). Restricting the study to Guatemala (gray polyhedron) will require to use species records only across this country (red and blue dots), resulting in a subset of occurrence records, and environmental tolerances, of the mosquito species (red and blue polyhedrons) **(A)**. Accounting for all the available data for the vector species, however, reveals that Guatemala only contains a portion of the species environmental tolerances **(B,C)**. As a result, when the study design is restricted to Guatemala as the study area **M**, and to its climate and environmental space, the model would yield in an incomplete reconstruction of the actual distributional potential of vector-borne diseases. Restricting the models to Guatemala would require other, fine-scale, landscape-level environmental variables. Otherwise, models will simply mirror the density of points. Models constructed using an envelope algorithm in NicheA (71).

"simulate" absence data to be able to use presence-absence algorithms. A common approach to simulate absence data is to generate random points across the study area. Presence-absence models that use simulated (i.e., fake) absence data during calibration are termed presence-background models. Presence-background algorithms thus use the same regression and classification algorithms used for presence-absence models, with the unique philosophical variation regarding the interpretation of absences vs. background points. Also, because the background corresponds to the study area, calibration of these algorithms is highly sensitive to variations in the extent of the study area extent selected.

Maxent is a popular ecological niche modeling algorithm based on logistic-like regressions comparing densities of occurrences (presences), densities of random points (background), and continuous environmental variables using diverse sets of parameters in the calibration process (47). Maxent protocols have been summarized in a series of software including *Wallace* (76), *dismo* (75), *ENMeval* (77), and *KUenm* (78) packages in R. Wallace is a user-friendly analytical environment to calibrate Maxent models, making it a good starting point for new users since it contains detailed instructions (76). Dismo provides less details regarding the different assumptions and complementary scientific literature,

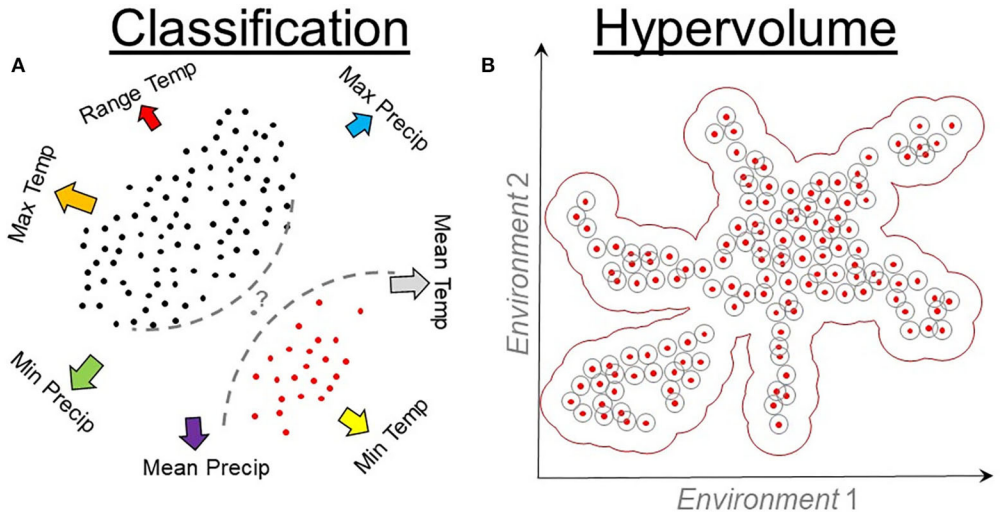


FIGURE 9 | Classification and hypervolume models. **(A)** Classification algorithms require environmental values where the species is present (red points) and absent (black points). Presence and absence data are linked to environmental values (arrows) to quantify the probability (question mark) of identifying a locality (gray point) as a suitable or unsuitable. Classification algorithms use these data to inform a series of rules (dashed lines) that vary among algorithms. Temp, temperature; Precip, precipitation. **(B)** Hypervolume algorithms quantify the density or cluster of presence records of the organisms in environmental dimensions. Hypervolumes measure the distance (gray cycles) among occurrences (red points) in an environmental space (arrows) to determine a best-fit model (red buffer).

but it is a good starting point for new users interested on modeling in programming environments (75). ENMeval is essentially the programming environment of Wallace and allows more detailed parameterization and evaluation of models (77). KUenm allows detailed, reproducible ecological niche models using Maxent and provides detailed model calibration and selection not available in the other packages (78), overcoming some of the perils of niche model applications for infectious diseases regarding differentiation between good and bad models (46). The KUenm package would be an ideal choice for advanced users since parameterization and installation would require advanced programming skills.

Presence-only algorithms focus solely on the environmental values linked to each occurrence record for calibration. As a result, calibration of these modes is insensitive to changes in the extent of the study area. Classic presence-only methods include environmental envelopes, which are ellipsoids, squares, or convex-hull that surround the occurrences in an environmental space (**Figure 9**), with algorithms that include Bioclim (75) and NicheA (71). Emerging presence-only methods include hypervolumes estimated using estimators of density (79) and cluster of occurrences in the environmental space (80). Protocols for hypervolume estimations have been described elsewhere (34, 74), and their use is expected to become common for $\mathbf{N}_R$ estimations due to the automatization of their workflows and computational optimization.

ECOLOGICAL NICHE MODELING AND CLIMATE CHANGE

A key set of questions in spatial epidemiology relates to effects of global change on the geographic distribution of infectious diseases and the potential of disease reservoirs or vectors to respond to such changes (81). Global change includes climate and land cover changes and the accelerated introduction of invasive species (82–84). A recent assessment proposed that catastrophic climate change effects will be perceived with even a 1.5°C annual mean temperature increase in the coming decades (85).

Ongoing climate change trends have been defined as human-induced, with unprecedented effects on biodiversity, impacting many organisms involved in disease transmission cycles (86). Climate change in the Anthropocene is generating geographic (87) and elevational (88) shifts of biodiversity, including organisms involved in disease transmission (89). Climate change is expected to produce bigger and more frequent weather events and wildfires (90, 91) and reductions of crop yields (92, 93), which together could generate ecological imbalance facilitating pathogen spillover (94). Understanding climate effects on directly transmitted diseases, however, remains in its infancy. Ecological niche models are a promising tool to help anticipate likely responses of disease systems to climate change. Recent assessments of vector-borne diseases have challenged paradigms related to climate and infectious diseases (95–97).

A recently published meta-analysis demonstrated that many popular algorithms for ecological niche modeling generally overestimate organisms' ability for adaptation to changing environments (98). The best forecasts should come from analyses of extensive data with simple algorithms (21, 99). That is, robust models require abundant, high-quality input occurrence data; these data are generally limited in availability in developing countries, so research about global change effects on diseases may be biased to developed countries (100). Nevertheless, even when data limitations may exist, ecological niche models provide opportunities to understand how global change can affect

infectious diseases globally. Based on the observations described above, when limited data are available, the use of multiple algorithms could help to explain uncertainty in model estimates.

The present understanding of potential climate change effects on organisms is biased geographically to temperate-zone countries (101). Nevertheless, tropical countries already show considerable climate change manifested in just the last three decades. Hence, the tropics represent an important priority for global change disease ecology research in view of their considerable research gaps, their role in modulating global climate, the need to understand organisms' responses to environmental change beyond temperate areas, and the need to assess niche evolution empirically in more rigorous analyses (102).

THE PARADOX OF DIRECTLY TRANSMITTED DISEASES

It is not surprising that ecological niche modeling applications in spatial epidemiology are biased toward vector-borne diseases. Data of disease vectors (e.g., fleas, mosquitoes, ticks) are broadly and openly available for many diseases and regions (103). Vectors are also highly responsive to changes in microclimate, with strong responses in their abundance, richness, distribution, and behavior linked to climate and landscape variation (104). As a consequence, ecological niche models of vector species provide good proxies of potential distributions of vector-borne diseases.

Environmentally transmitted diseases, such as anthrax, leptospirosis, and histoplasmosis, can also be studied using ecological niche modeling. Key components of the models include variables resembling the environmental drivers of parasite and pathogen persistence in the environment (e.g., humidity, temperature, and soil pH). When such variables are not available, some proxies could be used with their respective caveats.

Ecological niche models have many advantages compared with other disease modeling approaches, especially with regard to the biological bases that support the use of environmental drivers to map disease distributions. Nevertheless, ecological niche models are not suitable for the study of many disease systems, especially for studies aiming to understand direct transmission between individuals or populations. In such situations, other modeling approaches could be more appropriated (e.g., compartmental models). Similarly, ecological niche modeling may be a perilous modeling framework to use for animal disease systems where the environmental conditions are less important for transmission compared with animal density or human behavior.

Directly-transmitted diseases are more challenging to map based on environmental conditions. Many fine-scale factors (e.g., host density, age, immune status) shaping direct disease transmission may be required for correct reconstruction of transmission, but variables of such factors are generally not available. When the directly-transmitted disease includes an animal reservoir (e.g., wildlife), ecological niche models can focus on such species for the reconstruction of likely areas of transmission.

Ecological niche modeling of directly-transmitted animal diseases are a "dark side" that many veterinary epidemiologists avoid. Limited data of crucial factors associated with transmission and potential economic and ethical implications generally reduce explorations of directly-transmitted animal diseases. For example, the porcine reproductive and respiratory syndrome (PRRS, caused by a virus from the family Arteriviridae) affects the pork industry so that understanding and anticipating its distribution may have enormous benefits for its control and prevention in pig farms. Nevertheless, intensive farms may have controlled environmental conditions, so that the environmental conditions inside the pig farms may not reflect the surrounding climatic conditions. Thus, ecological niche modeling of PRRS risk based solely on the surrounding climate of farms is analytically and computationally feasible [e.g., (105, 106)], but such models will provide an erroneous signals of the environmental conditions suitable for transmission. That is, even when one can model linkages between climate and reports of directly-transmitted diseases, such models could be incomplete, biased, or misleading, and local factors may be more important (107). Paradoxically, models of directly-transmitted diseases are still popular.

CONCLUSION

Spatial epidemiology of animal diseases seems to be dominated by local-level studies (37). Thus, ecological niche modeling approaches provide an opportunity to reconstruct environmental conditions suitable for diverse animal diseases to identify areas where transmission is expected. Since disease systems need at least two organisms interacting (host and pathogen), biotic interactions may lie at the core of the pathogen's ecological niche, and neglecting interacting organisms in pathogen dynamics (i.e., maintenance, reproduction, transmission, and spread) may limit the success of forecasts. Pathogen transmission is strongly influenced by fine-scale interactions among infected and susceptible hosts, which can be further affected by host behavior and pathogen demography/transmission. Given the complexity of these interactions, traditional single-species ecological niche modeling approaches could fail to predict disease distributions and transmission risk accurately and protocols need to be revised with caution.

A new challenge in veterinary epidemiology is to avoid falling behind advances that distributional ecology offers in terms of theory and methods to map parasites, pathogens, vectors, and reservoir. This overview is by no means a detailed summary of all the advances in the field of ecological niche modeling. Instead, this review provides a brief introduction to the field facilitating a more effective use of the comprehensive ecological niche modeling courses freely available (e.g., https://www.youtube.com/watch?v=vj8qTo56rPA&ab_channel=A.TownsendPeterson) (108). Veterinary epidemiology needs more ecology, and ecologists modeling disease distributions need to incorporate health professionals for sound and biologically realistic model interpretations (15). Veterinary epidemiologists may find ecological niche modeling useful for disease control efforts, especially for infectious diseases with vectors or wildlife reservoirs. The limited presence of

epidemiologists and disease ecologists in the ecological niche modeling community increases the risk of inaccurate and misleading forecasting of infectious diseases of questionable quality and usefulness for stakeholders [e.g., (109)]. The broadly available epidemiological data, collected systematically from humans, animals, and plants, can help to advance the study of disease transmission. The comprehensive understanding of disease systems by veterinarians provides unique opportunities for their active participation in the field of spatial epidemiology. Nevertheless, mature and ethical ecological niche modeling applications for disease mapping would require familiarity with classic ecological theory.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was supported by Virginia Tech Startup Funds for LE.

ACKNOWLEDGMENTS

The author thanks A. T. Peterson for his contributions to the field in the form of online and in-person training, books, and manuscripts that triggered the main ideas of the new modeling framework. Paige Van de Vuurst, Steven. N. Winter, Diego Soler-Tovar, Natalie Brown, and Emily Hardgrove provided helpful comments in a final version of the manuscript. Thanks also to Jaber Amine Belkhiria, Jörn Gethmann, and Gustavo Machado for their evaluation and suggestions. Thanks to Salome Dürr, Victoria J. Brookes, and Andres M. Perez for the invitation to contribute to this special issue.

REFERENCES

1. Pfeiffer DU, Robinson TP, Stevenson M, Stevens KB, Rogers DJ, Clements ACA. *Spatial Analysis in Epidemiology*. Oxford: Oxford University Press (2008). doi: 10.1093/acprof:oso/9780198509882.001.0001
2. Ostfeld RS, Glass GE, Keesing F. Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends Ecol Evol*. (2005) 20:328–36. doi: 10.1016/j.tree.2005.03.009
3. Eisen RJ, Lane RS, Fritz CL, Eisen L. Spatial patterns of lyme disease risk in California based on disease incidence data and modeling of vector-tick exposure. *Am J Trop Med Hyg*. (2006) 75:669–76. doi: 10.4269/ajtmh.2006.75.669
4. Lash RR, Carroll DS, Hughes CM, Nakazawa Y, Karem K, Damon IK, et al. Effects of georeferencing effort on mapping monkeypox case distributions and transmission risk. *Int J Health Geogr*. (2012) 11:23. doi: 10.1186/1476-072X-11-23
5. Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, et al. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE*. (2017) 12:e0169748. doi: 10.1371/journal.pone.0169748
6. Horning N, Robinson J, Sterling E, Turner W, Spector S. *Remote Sensing for Ecology and Conservation*. New York, NY: Oxford University Press (2010).
7. Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, et al. Climatologies at high resolution for the earth's land surface areas. *Sci Data*. (2017) 4:1–20. doi: 10.1038/sdata.2017.122
8. Vega GC, Pertierra LR, Olalla-Táraga MÁ. MERRAclim, a high-resolution global dataset of remotely sensed bioclimatic variables for ecological modelling. *Sci Data*. (2019) 4:170078. doi: 10.1038/sdata.2017.78
9. Amatulli G, Domisch S, Tuanmu M-N, Parmentier B, Ranipeta A, Malczyk J, et al. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Sci Data*. (2018) 5:180040. doi: 10.1038/sdata.2018.40
10. Benavides JA, Valderrama W, Streicker DG. Spatial expansions and travelling waves of rabies in vampire bats. *Proc R Soc B*. (2016) 283:20160328. doi: 10.1098/rspb.2016.0328
11. Hutter SE, Brugger K, Sancho Vargas VH, González R, Aguilar O, León B, et al. Rabies in Costa Rica: documentation of the surveillance program and the endemic situation from 1985 to 2014. *Vector-Borne Zoonotic Dis*. (2016) 16:334–41. doi: 10.1089/vbz.2015.1906
12. Piaggio AJ, Russell AL, Osorio IA, Jiménez Ramírez A, Fischer JW, Neuwald JL, et al. Genetic demography at the leading edge of the distribution of a rabies virus vector. *Ecol Evol*. (2017) 7:5343–51. doi: 10.1002/ece3.3087
13. Soberón J. The relationship between use and suitability of resources and its consequences to insect population size. *Am Nat*. (1986) 127:338–57. doi: 10.1086/284488
14. Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, Nakamura M, et al. *Ecological Niches and Geographic Distributions*. Princeton: Princeton University Press (2011).
15. Peterson AT. *Mapping Disease Transmission Risk*. Baltimore: Johns Hopkins University Press. (2014).
16. Araújo MB, Anderson RP, Barbosa AM, Beale CM, Dormann CF, Early R, et al. Standards for distribution models in biodiversity assessments. *Sci Adv*. (2019) 5:1–12. doi: 10.1126/sciadv.aat4858
17. Cuervo-Robayo AP, Escobar LE, Osorio-Olvera LA, Nori J, Varela S, Martinez-Meyer E, et al. Introduccion a los análisis espaciales con enfasis en los modelos de nicho ecologico. *Biodivers Inform*. (2017) 12:45–57. doi: 10.17161/bi.v12i0.6507
18. Peterson AT, Anderson RP, Cobos ME, Cuahutle M, Cuervo-robayo AP, Escobar LE, et al. Curso modelado de nicho ecológico, versión 1.0. *Biodivers Informatics*. (2019) 14:1–7. doi: 10.17161/bi.v14i0.8189
19. Burnham KP, Anderson DR, Huyvaert KP. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav Ecol Sociobiol*. (2011) 65:23–35. doi: 10.1007/s00265-010-1029-6
20. Gotelli NJ, Ellison AM. *A Primer of Ecological Statistics*. 2nd ed. Sunderland: Sinauer Associates (2013).
21. Escobar LE, Qiao H, Cabello J, Peterson AT. Ecological niche modeling re-examined: A case study with the Darwin's fox. *Ecol Evol*. (2018) 8:4757–70. doi: 10.1002/ece3.4014
22. Dicko AH, Lancelot R, Seck MT, Guerrini L, Sall B, Lo M, et al. Using species distribution models to optimize vector control in the framework of the tsetse eradication campaign in Senegal. *Proc Natl Acad Sci USA*. (2014) 111:10149–54. doi: 10.1073/pnas.1407773111
23. Birch LC. Experimental background to the study of the distribution and abundance of insects: III. The relation between innate capacity for increase and survival of different species of beetles living together on the same food. *Evolution*. (1953) 7:136–44. doi: 10.1111/j.1558-5646.1953.tb00072.x
24. Haefner PA. Temperature and salinity tolerance of the Sand shrimp, *Crangon septemspinosa* Say. *Physiol Zool*. (1969) 42:388–97. doi: 10.1086/physzool.42.4.30155510
25. Maguire BJ. Niche response structure and the analytical potential of its relationships to the habitat. *Am Nat*. (1973) 107:213–46. doi: 10.1086/282827
26. Brown J. On the relationship between abundance and distribution of species. *Am Nat*. (1984) 124:255–79. doi: 10.1086/284267
27. Hooper HL, Connon R, Callaghan A, Fryer G, Yarwood-Buchanan S, Biggs J, et al. The ecological niche of *Daphnia magna* characterized using population growth rate. *Ecology*. (2008) 89:1015–22. doi: 10.1890/07-0559.1
28. Jiménez L, Soberón J, Christen JA, Soto D. On the problem of modeling a fundamental niche from occurrence data. *Ecol Modell*. (2019) 397:74–83. doi: 10.1016/j.ecolmodel.2019.01.020

29. Soberón J, Nakamura M. Niches and distributional areas: Concepts, methods, and assumptions. *Proc Natl Acad Sci USA*. (2009) 106:19644–50. doi: 10.1073/pnas.0901637106
30. Perkins TA, Metcalf CJE, Grenfell BT, Tatem AJ. Estimating drivers of autochthonous transmission of chikungunya virus in its invasion of the Americas. *PLoS Curr Outbreaks*. (2015) 1:1–19. doi: 10.1371/currents.outbreaks.a4c7b6ac10e0420b1788c9767946d1fc
31. Pritzkow S, Morales R, Moda F, Khan U, Telling GC, Hoover E, et al. Grass plants bind, retain, uptake, and transport infectious prions. *Cell Rep*. (2015) 11:1168–75. doi: 10.1016/j.celrep.2015.04.036
32. Pritzkow S, Morales R, Lyon A, Concha-Marambio L, Urayama A, Soto C. Efficient prion disease transmission through common environmental materials. *J Biol Chem*. (2018) 293:3363–73. doi: 10.1074/jbc.M117.810747
33. Pettorelli N. *The Normalized Difference Vegetation Index*. Oxford: Oxford University Press. (2013). doi: 10.1093/acprof:osobl/9780199693160.001.0001
34. Escobar LE, Qiao H, Lee C, Phelps NBD. Novel methods in disease biogeography: a case study with heterosporosis. *Front Vet Sci*. (2017) 4:105. doi: 10.3389/fvets.2017.00105
35. Owens HL, Campbell LP, Dornak LL, Saupe EE, Barve N, Soberón J, et al. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecol Modell*. (2013) 263:10–8. doi: 10.1016/j.ecolmodel.2013.04.011
36. Levin SA. The problem of pattern and scale in ecology: The Robert H. *MacArthur Award Lect Ecol*. (1992) 73:1943–67. doi: 10.2307/1941447
37. Winter SN, Escobar LE. Chronic wasting disease modeling: An overview. (in press). *J Wildl Dis.*. doi: 10.7589/2019-08-213
38. Zimmermann NE, Yoccoz NG, Edwards TC, Meier ES, Thuiller W, Guisan A, et al. Climatic extremes improve predictions of spatial patterns of tree species. *Proc Natl Acad Sci USA*. (2009) 106:19723–8. doi: 10.1073/pnas.0901643106
39. Kellermann V, Overgaard J, Hoffmann AA, Flojgaard C, Svenning J-C, Loeschcke V. Upper thermal limits of *Drosophila* are linked to species distributions and strongly constrained phylogenetically. *Proc Natl Acad Sci USA*. (2012) 109:16228–33. doi: 10.1073/pnas.1207553109
40. Johnson EE, Escobar LE, Zambrana-Torrelio C. An ecological framework for modeling the geography of disease transmission. *Trends Ecol Evol*. (2019) 34:655–68. doi: 10.1016/j.tree.2019.03.004
41. Peterson AT, Papeş M, Soberón J. Mechanistic and correlative models of ecological niches. *Eur J Ecol*. (2016) 1:28–38. doi: 10.1515/eje-2015-0014
42. Taylor RA, Ryan SJ, Lippi CA, Hall DG, Narouei-Khandan HA, Rohr JR, et al. Predicting the fundamental thermal niche of crop pests and diseases in a changing world: a case study on citrus greening. *J Appl Ecol*. (2019) 56:2057–68. doi: 10.1111/1365-2664.13455
43. Grinnell J. The niche-relationships of the California Thrasher. *Auk*. (1917) 34:427–33. doi: 10.2307/4072271
44. Maher SP, Ellis C, Gage KL, Enscore RE, Peterson AT. Range-wide determinants of plague distribution in North America. *Am J Trop Med Hygene*. (2010) 83:736–42. doi: 10.4269/ajtmh.2010.10-0042
45. Anderson RP. When and how should biotic interactions be considered in models of species niches and distributions? *J Biogeogr*. (2017) 44:8–17. doi: 10.1111/jbi.12825
46. Escobar LE, Craft ME. Advances and limitations of disease biogeography using ecological niche modeling. *Front Microbiol*. (2016) 7:1174. doi: 10.3389/fmicb.2016.01174
47. Merow C, Smith MJ, Silander JA. A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*. (2013) 36:1058–69. doi: 10.1111/j.1600-0587.2013.07872.x
48. Joppa LN, McInerny G, Harper R, Salido L, Takeda K, O'Hara K, et al. Troubling trends in scientific software use. *Science*. (2013) 340:814–5. doi: 10.1126/science.1231535
49. Anderson RP. Modeling niches and distributions: It's not just "click, click, click." *Biogeografía*. (2015) 8:11—27. Available online at: https://2278aec0-37af-4634-a250-8bb191f1aab7.filesusr.com/ugd/e41566_e8acb6f9c20c44fa9cd729161582857d.pdf
50. Peterson AT, Nakazawa Y. Environmental data sets matter in ecological niche modelling: An example with *Solenopsis invicta* and *Solenopsis richteri*. *Glob Ecol Biogeogr*. (2007) 17:135–44. doi: 10.1111/j.1466-8238.2007.00347.x
51. Escobar LE, Peterson AT. Spatial epidemiology of bat-borne rabies in Colombia. *Pan Am J Public Heal*. (2013) 34:135–6. Available online at: https://iris.paho.org/handle/10665.2/9141
52. Escobar LE. Modelos de nicho ecológico en salud pública: Cinco preguntas cruciales. *Pan Am J Public Heal*. (2016) 40:98. Available online at: https://iris.paho.org/handle/10665.2/31163
53. Cobos ME, Jiménez L, Nuñez-Penichet C, Romero-Alvarez D, Simoes M. Sample data and training modules for cleaning biodiversity information. *Biodivers Informatics*. (2018) 13:49–50. doi: 10.17161/bi.v13i0.7600
54. Escobar LE, Lira-Noriega A, Medina-Vogel G, Peterson AT. Potential for spread of white-nose fungus (*Pseudogymnoascus destructans*) in the Americas: Using Maxent and NicheA to assure strict model transference. *Geospat Health*. (2014) 11:221–9. doi: 10.4081/gh.2014.19
55. Pavlovsky EN. *Natural Nidality of Transmissible Diseases, with Special Reference to the Landscape Epidemiology of Zooanthroponoses*. Urbana: University of Illinois Press (1966).
56. Soberón J. Grinnellian and Eltonian niches and geographic distributions of species. *Ecol Lett*. (2007) 10:1115–23. doi: 10.1111/j.1461-0248.2007.01107.x
57. Schmidt KKA, Ostfeld RS. Biodiversity and the dilution effect in disease ecology. *Ecology*. (2001) 82:609–19. doi: 10.1890/0012-9658(2001)082[0609:BATDEI]2.0.CO;2
58. Food and Agriculture Organization. *Livestock Densities*. (2014). Available online at: http://www.fao.org/ag/againfo/resources/en/glw/glw_dens.html (accessed October 10, 2017).
59. Escobar LE, Awan MN, Qiao H. Anthropogenic disturbance and habitat loss for the red-listed Asiatic black bear (*Ursus thibetanus*): Using ecological niche modeling and nighttime light satellite imagery. *Biol Conserv*. (2015) 191:400–7. doi: 10.1016/j.biocon.2015.06.040
60. Escobar LE, Moen R, Craft ME, VanderWaal KL. Mapping parasite transmission risk from white-tailed deer to a declining moose population. (2019) 65:60. *Eur J Wildl Res*. doi: 10.1007/s10344-019-1297-z
61. Global Roads Inventory Project Transportation 4.0. Roads (2017). http://www.naturalearthdata.com/downloads/10m-cultural-vectors/roads/ (accessed July 11, 2017).
62. Meijer JR, Huijbregts MAJ, Schotten KCGJ, Schipper AM. Global patterns of current and future road infrastructure. *Environ Res Lett*. (2018) 13:064006. doi: 10.1088/1748-9326/aabd42
63. Gomes MN, Vieira Monteiro AM, Lewis N, Gonçalves CA, de Souza Nogueira Filho V. Landscape risk factors for attacks of vampire bats on cattle in Sào Paulo, Brazil. *Prev Vet Med*. (2010) 93:139–46. doi: 10.1016/j.prevetmed.2009.10.006
64. Braga GB, Grisi-Filho JHH, Leite BM, de Sena EF, Dias RA. Predictive qualitative risk model of bovine rabies occurrence in Brazil. *Prev Vet Med*. (2014) 113:536–46. doi: 10.1016/j.prevetmed.2013.12.011
65. Carrera JE, Bentz AB, Camus MS, Chizhikov VE, Recuenco SE, Czirják GÁ, et al. Livestock abundance predicts vampire bat demography, immune profiles and bacterial infection risk. *Philos Trans R Soc B*. (2018) 373:20170089. doi: 10.1098/rstb.2017.0089
66. Botto Nuñez G, Becker DJ, Plowright RK. The emergence of vampire bat rabies in Uruguay within a historical context. *Epidemiol Infect*. (2019) 147:e180. doi: 10.1017/S0950268819000682
67. Orlando SA, Panchana VF, Calderón JL, Muñoz OS, Campos DN, Torres-Lasso PR, et al. Risk factors associated with attacks of hematophagous bats (*Desmodus rotundus*) on cattle in Ecuador. *Vector-Borne Zoonotic Dis*. (2019) 19:407–13. doi: 10.1089/vbz.2017.2247
68. Varela S, Anderson RP, García-Valdés R, Fernández-González F. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography*. (2019) 37:1084–91. doi: 10.1111/j.1600-0587.2013.00441.x
69. Atauchi PJ, Peterson AT, Flanagan J. Species distribution models for Peruvian plantcutter improve with consideration of biotic interactions. *J Avian Biol*. (2018) 49:1–8. doi: 10.1111/jav.01617
70. Soberón J, Peterson AT. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodivers Informatics*. (2005) 2:1–10. doi: 10.17161/bi.v2i0.4
71. Qiao H, Peterson AT, Campbell LP, Soberón J, Ji L, Escobar LE. NicheA: Creating virtual species and ecological niches in multivariate environmental scenarios. *Ecography*. (2016) 39:805–13. doi: 10.1111/ecog.01961

72. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *J Anim Ecol*. (2008) 77:802–13. doi: 10.1111/j.1365-2656.2008.01390.x
73. Drake JM. Range bagging: A new method for ecological niche modelling from presence-only data. *J R Soc Interface*. (2015) 12:20150086. doi: 10.1098/rsif.2015.0086
74. Blonder B, Morrow CB, Maitner B, Harris DJ, Lamanna C, Violle C, et al. New approaches for delineating n-dimensional hypervolumes. *Methods Ecol Evol*. (2018) 9:305–19. doi: 10.1111/2041-210X.12865
75. Hijmans RJ, Elith, J. *Species Distribution Modeling*. (2020). RSpatial Available online at: https://rspatial.org/raster/sdm/index.html (accessed June 24, 2020).
76. Kass JM, Vilela B, Aiello-Lammens ME, Muscarella R, Merow C, Anderson RP. Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. *Methods Ecol Evol*. (2018) 9:1151–6. doi: 10.1111/2041-210X.12945
77. Muscarella R, Galante PJ, Soley-Guardia M, Boria RA, Kass JM, Uriarte M, et al. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods Ecol Evol*. (2014) 5:1198–205. doi: 10.1111/2041-210X.12261
78. Cobos ME, Peterson AT, Barve N, Osorio-Olvera L. kuenm: A dynamic R package for detailed development of ecological niche models using Maxent. *PeerJ*. (2019) 7:e6281. doi: 10.7717/peerj.6281
79. Blonder B, Lamanna C, Violle C, Enquist BJ. The n-dimensional hypervolume. *Glob Ecol Biogeogr*. (2014) 23:595–609. doi: 10.1111/geb.12146
80. Qiao H, Lin C, Jiang Z, Ji L. Marble algorithm: A solution to estimating ecological niches from presence-only records. *Sci Rep*. (2015) 5:14232. doi: 10.1038/srep14232
81. Shuman EK. Global climate change and infectious diseases. *N Engl J Med*. (2010) 362:1061–3. doi: 10.1056/NEJMp0912931
82. Armesto JJ, Manuschevich D, Mora A, Smith-Ramirez C, Rozzi R, Abarzúa AM, et al. From the Holocene to the Anthropocene: A historical framework for land cover change in southwestern South America in the past 15,000 years. *Land Use Policy*. (2010) 27:148–60. doi: 10.1016/j.landusepol.2009.07.006
83. van Asselen S, Verburg PH. Land cover change or land-use intensification: Simulating land system change with a global-scale land change model. *Glob Chang Biol*. (2013) 19:3648–67. doi: 10.1111/gcb.12331
84. Brockmann D, Helbing D. The hidden geometry of complex, network-driven contagion phenomena. *Science*. (2013) 342:1337–42. doi: 10.1126/science.1245200
85. Intergovernmental Panel on Climate Change. *Global Warming 1.5 °C*. Incheon: WMO/UNEP (2018).
86. Manning MR, Edmonds J, Emori S, Grubler A, Hibbard K, Joos F, et al. Misrepresentation of the IPCC CO_2 emission scenarios. *Nat Geosci*. (2010) 3:376–7. doi: 10.1038/ngeo880
87. Iknayan KJ, Beissinger SR. Collapse of a desert bird community over the past century driven by climate change. *Proc Natl Acad Sci USA*. (2018) 115:8597–602. doi: 10.1073/pnas.1805123115
88. Morueta-Holme N, Engemann K, Sandoval-Acuña P, Jonas JD, Segnitz RM, Svenning J-C. Strong upslope shifts in Chimborazo's vegetation over two centuries since Humboldt. *Proc Natl Acad Sci USA*. (2015) 112:12741–5. doi: 10.1073/pnas.1509938112
89. Escobar LE, Romero-Alvarez DA, Leon R, Lepe-Lopez MA, Craft ME, Borbor-Cordoba MJ, et al. Declining prevalence of disease vectors under climate change. *Sci Rep*. (2016) 6:39150. doi: 10.1038/srep39150
90. Jolly WM, Cochrane MA, Freeborn PH, Holden ZA, Brown TJ, Williamson GJ, et al. Climate-induced variations in global wildfire danger from 1979 to 2013. *Nat Commun*. (2015) 6:7537. doi: 10.1038/ncomms8537
91. Timmermann A, An S-I, Kug J-S, Jin F-F, Cai W, Capotondi A, et al. El Niño–Southern Oscillation complexity. *Nature*. (2018) 559:535–45. doi: 10.1038/s41586-018-0252-6
92. Trnka M, Rötter RP, Ruiz-Ramos M, Kersebaum KC, Olesen JE, Žalud Z, et al. Adverse weather conditions for European wheat production will become more frequent with climate change. *Nat Clim Chang*. (2014) 4:637–43. doi: 10.1038/nclimate2242
93. Carleton TA. Crop-damaging temperatures increase suicide rates in India. *Proc Natl Acad Sci USA*. (2017) 114:8746–51. doi: 10.1073/pnas.1701354114
94. Allen T, Murray KA, Zambrana-Torrelio C, Morse SS, Rondinini C, Di Marco M, et al. Global hotspots and correlates of emerging zoonotic diseases. *Nat Commun*. (2017) 8:1124. doi: 10.1038/s41467-017-00923-8
95. Mordecai EA, Paaijmans KP, Johnson LR, Balzer C, Ben-Horin T, de Moor E, et al. Optimal temperature for malaria transmission is dramatically lower than previously predicted. *Ecol Lett*. (2013) 16:22–30. doi: 10.1111/ele.12015
96. Mordecai EA, Caldwell JM, Grossman MK, Lippi CA, Johnson LR, Neira M, et al. Thermal biology of mosquito-borne disease. *Ecol Lett*. (2019) 22:1690–708. doi: 10.1111/ele.13335
97. Ryan SJ, McNally A, Johnson LR, Mordecai EA, Ben-Horin T, Paaijmans KP, et al. Mapping physiological suitability limits for malaria in Africa under climate change. *Vector Borne Zoonotic Dis*. (2015) 15:718–25. doi: 10.1089/vbz.2015.1822
98. Qiao H, Feng X, Escobar LE, Peterson AT, Soberón J, Zhu G, et al. An evaluation of transferability of ecological niche models. *Ecography*. (2019) 42:521–34. doi: 10.1111/ecog.03986
99. Breiner FT, Guisan A, Bergamini A, Nobis MP. Overcoming limitations of modelling rare species by using ensembles of small models. *Methods Ecol Evol*. (2015) 6:1210–8. doi: 10.1111/2041-210X.12403
100. Peterson AT, Soberón J. Essential biodiversity variables are not global. *Biodivers Conserv*. (2018) 27:1277–88. doi: 10.1007/s10531-017-1479-5
101. Mundim FM, Bruna EM. Is there a temperate bias in our understanding of how climate change will alter plant-herbivore interactions? A meta-analysis of experimental studies. *Am Nat*. (2016) 188:S74–89. doi: 10.1086/687530
102. Williams JW, Jackson ST. Novel climates, no-analog communities, and ecological surprises. *Front Ecol Environ*. (2007) 5:475–82. doi: 10.1890/070037
103. GBIF. *Global Biodiversity Information Faclity*. (2015). Available online at: http://www.gbif.org/ (accessed May 5, 2015).
104. Chandrasegaran K, Lahondère C, Escobar LE, Vinauger C. Linking mosquito ecology, traits, behavior, and disease transmission. *Trends Parasitol*. (2020) 36:393–403. doi: 10.1016/j.pt.2020.02.001
105. Alkhamis MA, Arruda AG, Morrison RB, Perez AM. Novel approaches for spatial and molecular surveillance of Porcine Reproductive and Respiratory Syndrome virus (PRRSv) in the United States. *Sci Rep*. (2017) 7:1–14. doi: 10.1038/s41598-017-04628-2
106. Alkhamis MA, Arruda AG, Vilalta C, Morrison RB, Perez AM. Surveillance of porcine reproductive and respiratory syndrome virus in the United States using risk mapping and species distribution modeling. *Prev Vet Med*. (2018) 150:135–42. doi: 10.1016/j.prevetmed.2017.11.011
107. Carlson CJ, Chipperfield JD, Benito BM, Telford RJ, O'Hara RB. Species distribution models are inappropriate for COVID-19. *Nat Ecol Evol*. (2020) 4:770–1. doi: 10.1038/s41559-020-1212-8
108. Peterson AT, Ingenloff, K. Biodiversity Informatics Training Curriculum, version 1.2. *Biodivers Informatics*. (2016) 10:65–74. doi: 10.17161/bi.v11i0.5008
109. Coro G. A global-scale ecological niche model to predict SARS-CoV-2 coronavirus infection rate. *Ecol Modell*. (2020) 431:109187. doi: 10.1016/j.ecolmodel.2020.109187

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer GM declared a past co-authorship with the author LE to the handling editor.

frontiers in Veterinary Science

REVIEW
published: 04 November 2020
doi: 10.3389/fvets.2020.532763

Participatory Epidemiology: Principles, Practice, Utility, and Lessons Learnt

Robyn G. Alders[1,2,3,4,5]*, *Syed Noman Ali*[6], *Aluma Araba Ameri*[7], *Brigitte Bagnol*[1,2,8], *Tarni L. Cooper*[1,9,10,11], *Ahmad Gozali*[5], *M. M. Hidayat*[12], *Elpidius Rukambile*[13,14], *Johanna T. Wong*[3,14] *and Andrew Catley*[15]

[1] Kyeema Foundation, Brisbane, QLD, Australia, [2] Kyeema Foundation, Maputo, Mozambique, [3] Center for Universal Health, Chatham House, London, United Kingdom, [4] Development Policy Center, Australian National University, Canberra, NSW, Australia, [5] Food and Agriculture Organization of the United Nations Animal Health, Jakarta, Indonesia, [6] Livestock Department, Government of Sindh, Karachi, Pakistan, [7] Private Consultant, Juba, South Sudan, [8] Department of Anthropology, University of the Witwatersrand, Johannesburg, South Africa, [9] School of Veterinary Science, The University of Queensland, Brisbane, QLD, Australia, [10] International Livestock Research Institute, Hanoi, Vietnam, [11] Centre for Communication and Social Change, School of Communication and Arts, The University of Queensland, Brisbane, QLD, Australia, [12] Directorate General of Livestock and Animal Health Services, Ministry of Agriculture, Jakarta, Indonesia, [13] Tanzania Veterinary Laboratory Agency, Dar es Salaam, Tanzania, [14] Faculty of Science, School of Life and Environmental Sciences, University of Sydney, Sydney, NSW, Australia, [15] Friedman School of Nutrition Science and Policy at Tufts University, Feinstein International Center, Boston, MA, United States

OPEN ACCESS

Edited by:
Victoria J. Brookes, Charles Sturt University, Australia

Reviewed by:
Jaime Ricardo Romero, Inter-American Institute for Cooperation on Agriculture, Costa Rica
Anke Wiethoelter, The University of Melbourne, Australia

***Correspondence:**
Robyn G. Alders
robyna@kyeemafoundation.org

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *05 February 2020*
Accepted: *02 October 2020*
Published: *04 November 2020*

Citation:
Alders RG, Ali SN, Ameri AA, Bagnol B, Cooper TL, Gozali A, Hidayat MM, Rukambile E, Wong JT and Catley A (2020) Participatory Epidemiology: Principles, Practice, Utility, and Lessons Learnt. Front. Vet. Sci. 7:532763. doi: 10.3389/fvets.2020.532763

Participatory epidemiology (PE) evolved as a branch of veterinary epidemiology and has been largely employed for the control and early warning of infectious diseases within resource-limited settings. It was originally based on combining practitioner communication skills with participatory methods to facilitate the involvement of animal caretakers and owners (embracing their knowledge, experience, and motivations) in the identification and assessment of animal disease problems, including in the design, implementation, monitoring and evaluation of disease control programs, policies, and strategies. With the importance of understanding social perceptions and drivers receiving increasing recognition by epidemiologists, PE tools are being adapted for an increasingly wide range of settings and endeavors. More recently, PE tools have been adapted for use in food and nutrition security programs, One Health activities, wildlife disease surveillance and as part of mixed-methods research across a range of socio-economic settings. This review describes the evolution of PE (in relation to veterinary epidemiology and briefly in relation to public health epidemiology), the underpinning philosophy and principles essential to its effective application and the importance of gender-sensitive approaches and data triangulation, including conventional confirmatory testing. The article also provides illustrative examples highlighting the diversity of approaches and applications of PE, hallmarks of successful PE initiatives and the lessons we can learn when these are missing. Finally, we look forward, describing the particular utility of PE for dealing with emerging infectious diseases, gaining attention of field-level cross-sector officials who can escalate concerns to a higher level and for continuing to raise the voices of those less-heard (such as women, minority groups, and remote communities with limited exposure to formal education) in defining the problems and planning activities that will likely impact directly on their well-being and livelihoods.

Keywords: participatory disease surveillance, medical anthropology, emerging infectious disease, One Health, participatory impact assessment

INTRODUCTION

Participatory epidemiology (PE) evolved as a branch of veterinary epidemiology and has been largely employed for the control and early warning of infectious diseases within resource-limited settings (1–4). These approaches and methods are derivatives of participatory appraisal and are useful in several conditions where the conventional epidemiological approaches do not provide the adequate level of understanding of the existing situation important for designing appropriate intervention. It was originally based on combining practitioner communication skills with participatory methods to facilitate the involvement of animal caretakers and owners (embracing their knowledge, experience, and motivations) in the identification and assessment of animal disease problems, including the design, implementation, monitoring and evaluation of disease control programs, policies, and strategies. This review describes the evolution of PE (in relation to veterinary epidemiology and briefly in relation to public health epidemiology), the underpinning philosophy and principles essential to its effective application, and highlights the importance of gender-sensitive approaches and data triangulation, including conventional confirmatory testing. It discusses the importance of understanding social perceptions and drivers, which is receiving increasing recognition by epidemiologists, and provides examples as to how PE tools are being adapted for an increasingly wide range of settings and endeavors, including: use in food and nutrition security programs (5–7); One Health activities (8); wildlife disease surveillance (9); gender analysis (10, 11); communication (12, 13); and for monitoring and evaluation (14).

HISTORY AND DEFINITION EVOLUTION

Paulo Freire (15) in "Pedagogy of the oppressed" advocated for a dialogue, and a participatory process for social transformation. By the late 1980s, there was a shift toward a more participatory approach to research, communication and extension services, particularly in the context of development activities. Consequently, participatory methodologies have been increasingly used in agricultural and livestock research development programs. Their use emerged in response to the failure of "normal" science to yield sustainable improvements to production and livelihoods in resource-limited, rural settings because of its inability to describe and intervene effectively in the complex and changing experiences of farmers and others involved in rural development (16). Early approaches were centralized and top-down (17). This top-down approach was unidirectional; initiated by the educated, expert, or intellectual (the "haves"), and directed toward the uneducated or ignorant (the "have nots"). This approach aimed to educate, convince or persuade individuals that their practices were wrong, and they should implement "modern" techniques. Chambers (18), publicized the idea of "putting the last first" and development organizations and extension services started to adopt some of these concepts. This led to demand-led extension, a process by which the information, advice and other extension services should be tailored to the expressed demands of the clients or users of the service (19–21). In participatory studies, knowledge is considered subjective and is generated through practical understanding of community practices (22). Subjective quality criteria are measured by the extent of individual's practical experience which leads to human improvement, hence the values of both the researcher and the participant are automatically brought into the research process.

Analysis of prior usage of participatory methodologies by Pretty (16) revealed at least seven different types of participation (**Table 1**) and lead to the recommendation that the

TABLE 1 | A typology of participation: how people participate in development programs and projects (16).

Passive participation	People participate by being told what is going to happen or has already happened. It is a unilateral announcement by an administration or project management without any listening to people's responses.
Participation in information giving	The information being shared belongs only to external professionals. People participate by answering questions posed by extractive researchers using questionnaire surveys or such similar approaches. People do not have the opportunity to influence proceedings, as the findings of the research are neither shared nor checked for accuracy.
Participation by consultation	People participate by being consulted, and external agents listen to views. These external agents define both problems and solutions and may modify these in the light of people's responses. Such a consultative process does not concede any share in decision making, and professionals are under no obligation to take on board people's views.
Participation for material benefits	People participate by providing resources such as labor, in return for food, cash or other material incentives. Much on-farm research falls in this category, as farmers provide the fields but are not involved in experimentation or the process of learning. It is very common to see this called participation, yet people have no stake in prolonging activities when incentives end.
Functional participation	People participate by forming groups to meet pre-determined objectives related to the project, which can involve the development or promotion of externally initiated social organization. Such involvement tends not to be at early stages of project cycles or planning, but rather after major decisions have already been made. These institutions tend to be dependent on external initiators and facilitators but may become self-dependent.
Interactive participation	People participate in joint analysis, which leads to action plans and the formation of new local institutions or the strengthening of existing ones. It tends to involve interdisciplinary methodologies that seek multiple objectives and make use of systematic and structured learning processes. These groups take control/ownership over local decisions, and so people have a stake in maintaining structures or practices.
Self-mobilization	People participate by taking initiatives independent of external institutions to change systems. Such self-initiated mobilization and collective action may or may not challenge existing inequitable distributions of wealth and power.

term "participation" should always come with the appropriate qualification. Our detailed review of PE was compiled with these different levels of participation in mind.

In 2000, Mariner and Paskin defined PE as "an emerging field that is based on the use of participatory techniques for the harvesting of qualitative epidemiological intelligence contained within community observations, existing veterinary knowledge and traditional oral history." Subsequently, Catley et al. (2) proposed a refined working definition of PE, i.e., "the systematic use of participatory approaches and methods to improve understanding of diseases and options for animal disease control." This working definition referred to "both a 'participatory approach' and 'participatory methods,' indicating that an understanding of both approach and methods are needed to define PE." Catley et al. (2) further proposed that "the term 'participatory' in PE is used to refer to the essential involvement of communities in defining and prioritizing veterinary-related problems, and in the development of solutions to service delivery, disease control, or surveillance.... use of the term PE that does not involve communities in these ways is considered to be a misnomer." In 2017, as part of a study of the major applications of PE in animal health, a modification of the Catley et al. (2) definition was proposed: "Participatory epidemiology is the systematic use of approaches and methods that facilitate the empowerment of people to identify and solve their health needs. It should promote the participation of people, leading to a shared learning environment that improves the understanding of their risk perception, health risks and options for surveillance, control, and health evaluation in populations. It should be conducted by professionals on equal partnership among all involved in the activity and with mutual respect and trust, ensuring acceptability and a sense of ownership" (3). This same study highlighted the utility of PE techniques in developing informed animal health policies by facilitating dialogue between communities and animal health officials in relation to disease prioritization. A 2020 review of PE disease control activities in pastoralist areas of Africa (4) examined the Allepuz et al. (3) modified definition by exploring the concept of empowerment within communities with significant socio-economic differentiation. Marked differences in wealth between households (4) and within households (23) have a significant effect on disease impacts and priorities and prevention and control preferences. Ensuring that PE techniques are applied through a gender-sensitive lens is crucial to achieving just and sustainable actions (2, 23).

The element of responsiveness or action combined with community engagement appears to set the PE employed within animal health apart from PE as employed within the public health arena. For example, "participatory" epidemiology has been used to refer to autonomous surveillance of social media for potential disease events (24). Bach et al. (25) conducted a review of the contribution of participatory research to epidemiology, emphasizing how participatory approaches can enhance common epidemiological approaches. The importance of the dissemination of findings was stressed by Bach et al. (25) but the need to actively work with communities to develop solutions appeared to be lacking in the review.

APPROACHES, METHODOLOGIES, AND TOOLS

Rapid rural appraisal (RRA) was a commonly employed, early approach to conducting a discrete study in one or more rural communities. These RRA studies were typically conducted within a week by a multidisciplinary team of researchers looking at a set of issues that were clearly defined by the study objectives (26). Participatory rural appraisal (PRA) subsequently emerged as an extended process that involves the collection of information and its eventual use by the community as it plans further activities (27). The aim of PRA is to stimulate a learning process and knowledge generation based on community members' experience to define priorities, and collect, analyse and interpret data (28–30). Participants are seen as the owners of the methods and outcomes of the appraisal. Participatory action research (PAR) goes a step further by utilizing the knowledge and understanding of community members as a point of reference to generate a participatory learning framework and actions. Research participants bridge the gap between the researcher and the researched by engaging in the data collection and scrutiny, and determination of the achievement trend of the research (22, 31). The ultimate goal of PAR is practical knowledge generation, making sure that the knowledge is made available and used for the transformation and empowerment of the individual participants and community at large (32). Participatory studies, including those beyond PE, deploy a wide range of techniques for data collection including but not limited to personal interviews, focus group discussions, observation, free listing, ranking, pair-wise ranking, causal flow analysis, open-ended stories, genograms, role playing, body mapping, and photo voice (30, 33, 34). The tools used for data collection in PRA and PAR should ensure gender inclusion and reduced gaps between the literate and illiterate to increase the chances of achieving equal access during information generation and sharing.

Table 2 summarizes the range of PE methodologies and tools that are now regularly in use in the field across a range of settings. Key to the successful use of these methods is an understanding of and commitment to: (i) the principles of adult learning (i.e., adult learners have different experiences, perceptions, problems and needs, and activities are more effective if trainers and PE practitioners understand how and why adults learn), (ii) triangulation (i.e., using more than one method to collect data on the same topic to verify findings, including multiple qualitative sources and participants, the use of secondary documentation, clinical examination, and laboratory diagnostic tests), and (iii) laboratory diagnostic support (i.e., in cases of livestock disease investigation, the use of PE tools needs to be accompanied by laboratory confirmation as it is not enough to rely on data collection using PE tools only).

A number of PE training documents are freely available online as are explanations of novel uses of these tools to tackle a range of animal, human, and One Health issues (**Table 3**).

TABLE 2 | An overview of the most commonly used PE methods and tools used to obtain specific information.

Method	Tools	Examples of data gathered
Informal interviewing (semi-structured)	Key informant interviews Focus-group discussions	Personal and group accounts of disease history and impacts Identification of important stakeholders
Ranking and scoring	Simple ranking Pair-wise ranking Proportional piling Matrix scoring Wealth ranking	Preferred types of livestock reared Relative livestock ownership Relative importance of livestock to livelihoods
Visualization	Participatory mapping Venn diagrams Seasonal calendars Timelines	Ecosystem boundaries and natural resources Veterinary services Seasonal variations in livestock disease Infrastructure Timeline of disease emergence and associated events
Direct observation	Transect walks Walking surveys	Infrastructure available Local environment Local living and working conditions Potential drivers of disease (such as water bodies, animal movements and interactions) Distance examination of animals for signs of disease
Participatory disease surveillance	The entire suite of participatory tools listed above applied to the disease of interest (usually based on syndromic diagnosis)	Information to develop a case definition Existence of or estimate of prevalence, incidence, morbidity and/or mortality of disease of interest

TABLE 3 | A compilation of PE methodology and training resources available free of charge online.

PE component	Potential applications	Source
Manual on participatory epidemiology	Action-oriented epidemiological intelligence collection and joint analysis	(35)
Participatory methodologies for use in pastoral areas	Disease surveillance in areas where animal healthcare and disease reporting systems are limited. To support the joint preparation of feasible and acceptable disease control strategies.	(1)
Participatory Epidemiology: a guide for trainers	Animal health surveys Problem analysis Disease detection Changing disease patterns Research	(36)
Participatory impact assessment	Participatory development of impact indicators by a range of stakeholders.	(37)
Participatory methodologies for family poultry production through a gender lens	To improve husbandry and biosecurity measures, and therefore health and production within small-scale chicken production systems for men and women farmers.	(38)
Trainer toolkit	A toolkit to assist in the implementation of introductory training programs in PE for adult students, including mid-career professionals.	(39)

IMPLEMENTATION EXPERIENCES AND LESSONS LEARNT FROM THE FIELD

The selection of case studies below provide an insight into the utilization of PE approaches within countries, initially in relation to animal disease prevention, then in relation to the linkages between animal disease and human food and nutrition security and finally, antimicrobial resistance. While far from an exhaustive list, the studies were selected to provide a diverse overview of geographical, cultural, disease, and methodological applications of PE over the last 20 years.

The first case study from Pakistan provides an overview of how the application of PE evolved over time and demonstrates how participatory epidemiology helped to shift the focus from the three diseases targeted by international agencies, i.e., rinderpest, foot and mouth disease (FMD) and *peste des petits ruminants* (PPR) to haemorrhagic septicaemia which was of greater concern to local farmers. The second case study from Sudan illustrates the variety of uses of PE and how it contributes to strengthen under-resourced health services. The Indonesian case study highlights the evolution of PE methods from an animal health focus to a broader One Health framework. Moving on to more recent project-specific examples with greater integration of One Health, the fourth case study from Tanzania connects participatory animal health to participatory nutritional security and food safety through a gender lens. In Timor-Leste, gender-sensitive participatory approaches were used to learn about animal disease, household food choices and food safety, while the case study from Uganda revealed how underlying causes of malnutrition were related to gender issues. Finally, the last case study from Vietnam provides insight into the use of participatory tools improve our understanding of and response to antimicrobial resistance (AMR).

Most of the case studies reflect the experiences of country nationals employing PE in support of national priorities. The case studies also emphasize the importance of employing PE techniques as part of a suite of activities that address the limitations of PE while also indicating how PE can contribute to multi-sectoral and interdisciplinary studies of complex systems.

Pakistan: From Global to Local Priorities

The Islamic Republic of Pakistan, situated in South Asia, is the world's fifth-most populous country with a population exceeding 212.2 million (40). The geography and climate of Pakistan are extremely diverse; it is divided into three major geographic areas: the northern highlands, the Indus River plain, and the Baluchistan Plateau. Correspondingly, the climate varies from tropical to temperate, with arid conditions in the coastal south. Rainfall varies greatly from year to year, and patterns of alternate

flooding and drought are common in the plains of Pakistan. Arable agriculture is mainly confined to the central fertile plain of the Indus River. Livestock production is a noteworthy section of agriculture primarily active in the arid and hyper-arid zones with restricted resources. Three systems of production systems are reported nationally according to the agroecological zone, i.e., nomadic, transhumant and stationary, or family business (41–43). Veterinary services in the remote areas of the country are poor and livestock owners mostly depend upon local herbal treatment practiced by families for decades.

During the second half of the twentieth Century, countless rural poverty alleviation programs that were developed and executed in the country, mainly in remote areas, unfortunately failed because of the gap between the farmers' views about their requirements and the understanding of the agencies that developed the programs (44). In the livestock sector, poor disease awareness, and reporting systems contributed to gaps in the design and implementation of animal disease control and eradication strategies as highlighted during the Global Rinderpest Eradication Program (45). The success stories of the participatory disease surveillance (PDS) active surveillance method employed in Africa (described in more detail in the case study on South Sudan below) prompted the Ministry of Food, Agriculture and Livestock, and Provincial Livestock Departments to introduce it into the country in support of transboundary animal disease (TAD) control. Participatory disease surveillance was implemented as a consultative process that proved to be valuable during the rinderpest eradication campaign from 1999 to 2007. Data obtained from PE was used to revise and improve rinderpest control methods and norms, both nationally and internationally (2, 46, 47). The PDS program greatly boosted the sensitivity of active clinical rinderpest surveillance and was pivotal to Pakistan's decision to declare provisional freedom from rinderpest to the OIE in January of 2003 (48). The integration of PDS with passive surveillance systems, based on reports from government and private veterinarians, enhanced their effectiveness by aggregating the number of cases detected for disease investigation and the timeliness of detection (47).

The occurrence of various important livestock diseases, particularly TADs such as Food and Mouth Disease (FMD; cattle and buffaloes) and *Peste des petits ruminants* (PPR) (sheep and goats) in the country were determined by applying different PDS tools. A full review of the data collected revealed that although FMD was the most prevalent disease, haemorrhagic septicaemia was considered the most important by farmers. Disease intelligence was gathered through various PE tools including visualization, scoring, and interview techniques (44, 46). Additional livestock health constraints documented during the field disease search program were mastitis, respiratory syndrome, intestinal parasite infestation, and buffalo pox. Gathering disease information through the application of participatory tools was a new approach in Pakistan. Initially, the majority of dairy farmers were hesitant about sharing their information and reluctant to actively participate in group discussions. Fortunately, as they came to understand that their indigenous knowledge was important and valued, it became relatively straight forward to obtain information pertinent to particular areas. The breadth and quality of data accrued through the application of PDS methodology has been valued by all livestock departments across the country. The estimation of disease prevalence and prioritization of their importance through PDS activities has helped to better plan and execute measures for the control/eradication of livestock diseases in different parts of the country. This approach was also found to be a practical option for obtaining reliable data that could be utilized by policy makers in their formulation of animal disease control and eradication in Pakistan (46). The most recent study was carried out in Tharparkar District of Sindh Province (44) in association with preventive vaccination against PPR disease.

Key lessons learnt to date in Pakistan, especially in relation to social behavior, include:

- Using a variety of exercises during interviews—such as scoring, mapping, and visualization—made it easier for farmers to share their point of view on various issues regarding livestock disease and the associated impact on their livelihoods;
- Some farmers hesitate to share information about infectious diseases in the presence of government veterinary staff;
- The PE approach was quite helpful when evaluating the disease situation in specific villages/areas. The interest of farmers/participants was very much evident during mapping, seasonal calendar and proportional piling exercises;
- Working with physical items that can be used to allocate preferences (e.g., stones/beans/seeds) during exercises was very effective in large groups and with key informants in rural areas. However, in peri-urban areas, livestock farmers preferred working with markers and charts;
- Through PDS activities, FMD, and PPR were found to be endemic throughout country. Farmers had been confusing PPR with contagious caprine pleuropneumonia and enterotoxaemia. Participatory disease surveillance teams confirmed PPR virus circulation serologically in villages of the country (46, 49);
- Foot and Mouth Disease and PPR were revealed, through participatory activities associated with the Rinderpest Eradication Program, to be causing socio-economic impacts that contributed to household poverty, malnutrition, starvation, and human health complications in rural areas of the country where mixed farming was common;
- Including female veterinarians in the PDS team was very successful in obtaining firsthand information from women who were directly involved in livestock management. Due to social restrictions male staff could not speak directly with women farmers; and
- Through the application of PDS tools (especially scoring and ranking tools), government veterinary services learnt that haemorrhagic septicaemia was of greater concern to farmers than the three diseases targeted by international agencies, i.e., rinderpest, FMD, and PPR.

South Sudan: Community Engagement Strengthens Effectiveness of Under-Resourced Health Services

The Republic of South Sudan, one of the world's newest countries, covers an area roughly the same as France, and is bordered by

Sudan, Central African Republic, Democratic Republic of Congo, Uganda, Kenya and Ethiopia. It has a variety of ecological zones, ranging from the flat savannah and flood plains around the Nile, and its tributaries, to the stony semi-arid region of the southeast to the rain forest of the undulating ironstone plateau of the west and south west. The climate fluctuates from very hot and dry in the dry season to hot and humid in the long rainy season when the low-lying areas are flooded, and every few years there are climatic extremes causing severe drought or floods (50).

In South Sudan, where population density is relatively low, infrastructure poor and ready access to human, and animal health services extremely limited in much of the country, a mix of consultative and interactive PE activities have played a vital role in disease control activities implemented through a One Health lens. "One Health" is the integrative effort of multiple disciplines working locally, nationally, and globally to attain optimal health for people, animals, and the environment (51). In the One Health space, consultative PE methodologies have been employed to conduct community-based surveillance and response systems for highly pathogenic avian influenza (HPAI) H5N1 from 2007 to 2009 [supported by USAID; (52)], and anthrax disease outbreak surveillance and control in humans and livestock in 2018 [including in South Sudanese refugees in Uganda; (53)]. The first wave of HPAI H5N1 outbreaks reported in poultry in Africa occurred in 2006, affecting eight African countries (Burkina Faso, Cameroon, Côte d'Ivoire, Djibouti, Egypt, Niger, Nigeria, and Sudan) in 2006 and three countries (Benin, Ghana, and Togo) in 2007 (54). A One Health approach was also adopted in South Sudan, involving veterinarians and human doctors, to conduct joint disease surveillance to investigate Rift Valley Fever (RVF) in the Lakes State. The approach resulted in the successful containment of RVF in livestock and human populations in the aforementioned state (55). Resources used to conduct interactive PE activities resulted not only in improved understanding of disease situations, they also simultaneously contributed to the development of collaborative approaches to disease surveillance and control.

Animal health studies utilizing PE to date have included applied research on a chronic wasting disease in cattle (called *liei* locally), impact assessment of community-based animal health projects, and the application of participatory disease searching during the rinderpest eradication program (36). Participatory epidemiology was a crucial component of rinderpest disease searching in 2002–2007, and also for FMD (56) in remote areas where classical veterinary surveillance activities would have been difficult to implement. In each case, the methods used to obtain information from stakeholders (including livestock owners, livestock traders, local authorities in government offices, veterinarians, Community Animal Health Workers, youth, women, and men) depended on the objective of the disease control activity. For example, for rinderpest disease eradication, consultative participatory disease search methodologies were used to locate rinderpest virus foci in villages where veterinary services were limited during the civil war (1983–2005) (50). Professionals in South Sudan have applied a wide range of PE tools, including semi-structured interviews, seasonal calendars, simple ranking, proportional piling (PP), PP for morbidity and mortality, timelines, and participatory disease searching. It has been noted that the practical value of PE in South Sudan demonstrates that it should be valued as an essential skill for field veterinarians and livestock officers, working for government or NGOs (57).

Examples and utility of PE tools that are frequently employed in South Sudan include:

- Participatory Mapping (PM) is used when consulting livestock herders regarding seasonal grazing patterns and this information helps in designing vaccination campaigns with livestock owners in a participatory manner. Participatory mapping is mostly done at the beginning of focus group discussions (FGDs) and key informant interviews (KIIs) as way to break the ice and allow free interaction with the participants. It is especially useful when PE team members are visiting for the first time and know little about the area and the community leaders. Mapping provides key information concerning resources available (water, rivers, hills, and pastures), distance covered searching for grazing land, identifying neighboring community, villages, infrastructure like market points, social centers, and proximity of government and private services to the livestock owners. It can reveal the livestock species in the grazing sites, wildlife species, insecure areas where livestock theft is common, and conflict amongst neighbors, for example due to scarcity of water and pasture. All of this valuable information can be obtained in 1 h and helps to break down barriers between visiting teams and key informants, local leaders, local authority, and community members;
- Simple Ranking (SR) and PP are easy to use with individual participants, KIIs and FGDs. They provide good information for planning and further research. A SR exercise uses objects or cards that can be easily placed in order of priority based on information provided by participants. A PP exercise is conducted using cards or objects to represent issues with participants placing counters on issues proportionally to the size of the problem represented. The bigger the pile against particular card or object, the larger the concerns of participants regarding that problem. Simply ranking and PP exercises were done separately for each gender (i.e., men and women). It was found that when combining men and women into one discussion group, men tended to dominate and push their opinions above those raised by women, impeding the process of building consensus concerning key information discussed during the PE activities. Separating groups by gender can facilitate an environment where women can comfortably share their opinions and ideas.

These three PE tools (i.e., PM, SR, and PP) facilitated consultation and interaction with participants and generated considerable amounts of information, with elaborate details frequently emerging that the PE team used to probe further to generate useful data for disease outbreak investigation or project design and implementation. For example, in August 2006 in Juba, a PE team, composed of mainly veterinary officers, used participatory mapping to identify where poultry were dying with simple ranking used to gauge disease morbidity and mortality

rates. These exercises were done with poultry owners who had reported sick chickens in Hai Jalaba. The sickness was perceived by livestock owners to be like Newcastle disease (ND). On the basis of the information provided by the owners the PE team suspected high pathogenic avian influenza (HPAI). As part of the triangulation process, samples were collected and dispatched to an OIE reference laboratory in the UK. Laboratory results confirmed the presence of HPAI H5N1 triggering the implementation of control activities.

Indonesia: From Participatory Animal Health to One Health

The Republic of Indonesia is a country in Southeast Asia and Oceania consisting of more than 17,000 islands (58). It is the world's largest island country and the 14th largest country by land area. With over 267 million people, it is the world's fourth most populous country as well as the most populous Muslim-majority country. Java, the world's most populous island, is home to more than half of the country's population. Indonesia's size, tropical climate, and archipelagic geography support one of the world's highest levels of biodiversity (59).

In 2005, Indonesia became one of the Asian epicenters for human and animal HPAI H5N1 infections during the global pandemic (60). The Indonesian Ministry of Agriculture (MoA) and Ministry of Health (MoH), together with the Coordinating Ministry of Human Development and Cultural Affairs (MoHDCA) worked with the Food and Agriculture Organization of the United Nations (FAO) to control the H5N1 outbreak and continue to work on a pilot research and development program to identify sustainable strategies for strengthening capacities for One Health-focused, effective and sustainable prevention and control of targeted zoonoses and emerging infectious diseases (EIDs).

The participatory disease surveillance and response (PDSR) program, developed to tackle HPAI H5N1 in Indonesia, was an evolution of the consultative PDS system employed during the rinderpest eradication program in African countries and Pakistan (60). The first stage of the PDSR project commenced in January 2006 and focused on the detection and control of HPAI (H5N1) by separate PDS and participatory disease response (PDR) teams, primarily in extensively raised poultry kept by households within village settings. Lessons learned during the first phase were used to strengthen disease management during the second phase of the project (with field implementation starting in May 2008) by adapting technical approaches to HPAI disease control, increasing functional participation of key stakeholders, including relevant district, provincial and central government agencies, and focusing on the community level. The PDSR project concluded in September 2015 with the end of the FAO ECTAD Avian Influenza prevention and control program in Indonesia. The PDSR program focused almost entirely on HPAI with little to no attention paid to other diseases of poultry (61) as external donor funding was largely driven by the public health desire to prevent an avian influenza pandemic in humans.

Subsequently, the Indonesian Ministry of Environment and Forestry joined with the MoA, MoH, and MoHDCA in collaboration with the FAO to develop sustainable strategies for strengthening One Health-focused, effective and sustainable prevention and control of targeted endemic zoonoses (i.e., anthrax, avian influenza, and rabies) and EIDs (62). Commencing in 2017, an agreement between four collaborating ministries was signed and four pilot districts were selected covering different agro-ecological zones. Master trainers across human, animal, and wildlife health were nominated by their district agencies and together with a central training team, PE tools such as PM and PDS were adapted to support One Health field investigations of reports of zoonotic disease. The PE tools proved readily adaptable for use by wildlife health officers, many of whom did not have a background in veterinary science. Notable success was achieved in relation to the prevention and control of rabies, which is endemic in many provinces and is the most commonly reported zoonotic disease. Prior to One Health PE training, 99% (1152/1155) of bite cases were reported via the human health system only. After 18 months, 50% (431/855) of reported cases were managed via a One Health integrated bite case management protocol. Integrated bite case management reports ($n = 431$) increased from 1% before training to 50% post-One Health PE training (8). Overall, the Zoonoses Prevention and Control programme in Indonesia effectively incorporates the One Health approach within its multisectoral field operations and associated multisectoral communication and information sharing platforms. This programme provides a template for the operationalization of participatory One Health approaches in Indonesia and beyond. Moreover, through the involvement of economists in the One Health team, the programme was able to demonstrate that it was highly-cost effective, generating 6.6-14.4 USD in benefits per dollar invested (63). These findings together with effective intersectoral collaboration and positive feedback from communities lay the foundation for the development of the National Master Plan for the eradication of rabies using a One Health framework (64).

A One Health PE approach integrating human, animal, and wildlife health provides an opportunity to detect novel pathogens prior to their transmission to humans. The focus on existing zoonotic disease accommodated the immediate priorities of communities and frontline officers while simultaneously building effective disease prevention and control systems (8).

Tanzania: From Participatory Animal Health to Participatory Nutritional Security and Food Safety

The United Republic of Tanzania is located in the eastern part of Africa and is about one tenth the size of the USA. Tanzania is bordered by Kenya and Uganda in the north; Burundi, Rwanda in the northwest; Democratic Republic of the Congo in the west; Malawi, and Zambia in the southwest; Mozambique in the south and Indian Ocean. Among the members of East Africa community, Tanzania has the largest population estimated at 58,552,845 and the lowest population density with almost one third of the population living in urban areas (65). From 1991 to 2015, the country has achieved significant decreases in stunting in children under 5 years of age from 50 to 35% and 22 to

12% of severe stunting in the same period (66). The Tanzania Demographic and Health Survey 2015–16 dataset indicates a prevalence of diarrhea in children under five of 12% (67), while diarhoea-specific mortality in the same age declined by 89% between 1980 and 2015 (68). Despite decreases in stunting and diarrhea-specific mortality, undernutrition and diarrhea in under five children are still important health problems and several efforts have been made through different platforms to overcome the problem including promoting animal and crop production and development of community-based educational packages.

An interdisciplinary and multi-sectoral team worked with local communities between 2014 and 2019 to strengthen traditional integrated livestock-crop systems in a semi-arid area of Central Tanzania with support from the Australian Center for International Agricultural Research (13, 69, 70). Representatives of the agriculture and health departments in Manyoni and Mpwapwa Districts were the key focal points of the project and were involved fully in the selection of participating wards. The village leaders from all three Wards were involved in the selection of men and women key informant interview (KII) and focus group discussion (FGD) participants and local project community workers. The community workers included village chicken vaccinators tasked with vaccinating household chickens during vaccination campaigns on a fee-for-service basis, the community assistants tasked with collecting fortnightly household data and enumerators who administered the questionnaire. Interventions targeted reduced mortality in extensively raised indigenous chickens through regular vaccination against Newcastle disease (ND) and constraints to the production and storage of nutrient-rich vegetables, grains, and pulses. Rural communities reliant on rain-fed crops often experience severe hunger periods immediately before the major harvesting season, when the previous year's stored grains have been exhausted or lost as a result of poor storage. Data was collected on human health and nutrition and household characteristics on a 6-monthly basis, livestock ownership on an annual basis, and chicken numbers and reports of diarrhea in children fortnightly as part of a cluster-randomized controlled trial involving children <24 months of age at the time of enrolment.

To facilitate the active engagement of all enrolled households, pictorial record charts were distributed at 4 monthly intervals, in the months of August and December in 2014 and April, August, and December in 2015. The aim was to document the consumption of poultry products over a period of 4 consecutive weeks (5). This research tool was developed by anthropologist B. Bagnol for use in communities with low levels of literacy. It was adapted from an approach used in reproductive health research in Tanzania and Uganda (71, 72) to enable the involvement of those without an understanding of written language. Black and white line drawings depicting a chicken, eggs, an infant, a pregnant woman, and a breastfeeding mother were presented in a table layout (**Figure 1**). In advance of each data collection period, the locally-selected Community Assistants were trained to instruct a household representative to use a mark to record any meal containing chicken or egg consumed by the enrolled child or by a pregnant or breastfeeding woman in their household if present. Each household was visited by a Community Assistants visited on a weekly basis to review the pictorial charts and assist participants in recording any incomplete data as required. Triangulation of data was achieved using data from the visual diaries, annual gender disaggregated focus group discussions (11, 13) and quantitative survey tools.

Food safety is increasingly being recognized as a key component of food and nutrition security (73). Epidemiological studies indicate a significant association between unhygienic food handling and occurrence of childhood diarrhea diseases which suggests food contamination can result in acute and/or chronic gastrointestinal infections (74, 75). diarhoeal diseases, which in most cases occurs as a result of consumption of contaminated food, are associated with high morbidity and mortality, especially in children <5 years of age in many low- and middle-income countries (LMICs) (76). Environmental microbes are important sources of food contamination and the routes in which these microbes enter the human food chain vary from one setting to another. Poor water supply, sanitation services and unhygienic practices accompanied by extensive animal keeping which favor human-animal proximity increases the risks of environmental- and animal-associated microbes to enter the human food chain. A qualitative study with 10 KIIs and 8 gender-segregated FDGs (four FGDs with women and four with men) with an average of 8 participants was conducted in resource-poor settings in central rural Tanzania to explore challenges associated with water supply, sanitation services, hygiene practices, and animal husbandry, seen to be important underlying factors related to childhood diarrhea (77, 78). Also, community knowledge and perceptions of the causes and occurrence of diarrhea in children was examined as understanding this is essential to designing effective prevention and control of childhood gastrointestinal infections.

While the overarching 5-year study sought to achieve interactive participation, the food safety study employed participation by consultation listening to the views of the study participants on the components being studied. The researcher defined the problem and guided the participants through the discussion by ensuring equal opportunity for all participants until contributions had been exhausted. The questionnaire survey revealed that households switch water sources between the dry and rainy seasons, especially in areas with public taps. The survey findings alone were not enough to explain the reason for this shift. By engaging with FGD participants, it became clear that a large proportion of the households use ground water (rivers, pond, and streams) during the rainy season because it is free and convenient; whereas during the dry season most households use public taps as their main source of water when accessible because other sources are no longer available. These findings are important as the National Water Policy of 2002 promotes the use of improved water sources, including the public taps, through user pay systems without adequate consideration of the impact of these costs on compliance. Through KIIs and FGDs it was clear that water shortage was a barrier to handwashing with soap and water, explaining that it is a difficult and expensive practice to maintain when water is scarce and or must be purchased.

Incorporation of participatory methods and ensuring community participation right from the inception stage of the project were significant contributing factors to obtaining

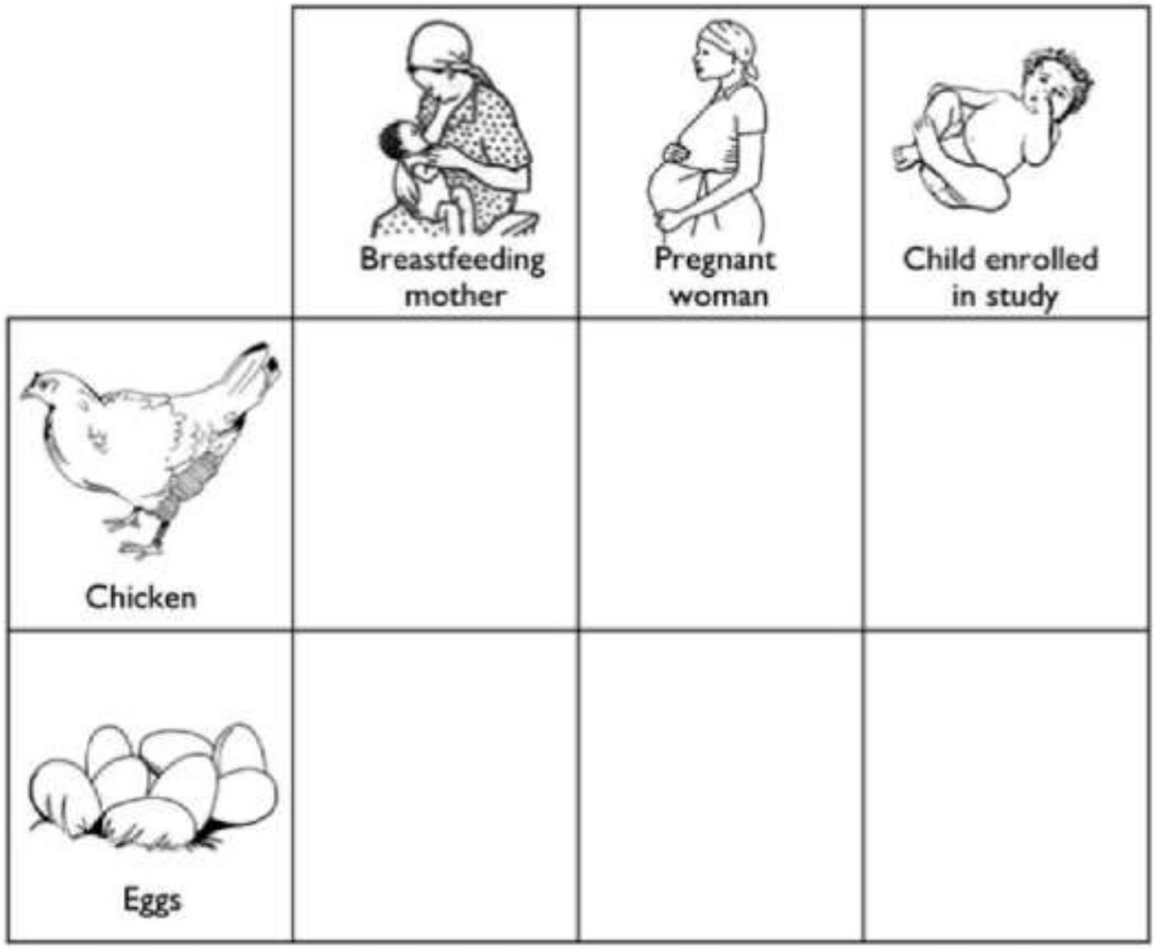


FIGURE 1 | A copy of the pictorial record chart (with English translations of Swahili text) for completion by a household representative, to record the consumption of poultry products by children enrolled in the study, and any pregnant or breastfeeding women within the same household [Source: (5)].

improved understanding of community perceptions and their decision-making processes. The KIIs and FGDs conducted to triangulate the data obtained by questionnaire survey provided an insight into the understanding of the community regarding key issues relating to the availability, suitability, and appropriate use of the water services available. The qualitative findings highlighted potential entry points for effective control of childhood gastrointestinal infections. The use of participatory methods and community engagement provided an insight into community perceptions regarding unhygienic practices and the effective use of available resources. Compiling and analyzing community perceptions is the key determinant for successful adoption of co-designed interventions.

Timor-Leste: Participatory Approaches to Learning About Animal Disease, Household Food Choices and Food Safety

Timor-Leste is a young, post-conflict country in Southeast Asia with a population of 1.27 million people in 2018 (79). Infrastructure is still rudimentary in many rural areas, and development is hindered by the challenging terrain and climatic conditions in much of the country (80, 81). Timor-Leste suffers high rates of child undernutrition, with 46% of children under five suffering from stunting in 2016, and children have low consumption of nutrient-rich animal-source food (82). In rural regions of Timor-Leste, heavy reliance on subsistence agriculture creates strong seasonal patterns in food availability and consumption (81). Household food insecurity exists when crop stores have been exhausted, and growing crops are not ready to be harvested (36). These patterns of crop availability, as well as seasonal foraging for wild vegetation, have been documented in parts of Timor-Leste (81, 83), however little data existed regarding the seasonality of animal-source food consumption, or consumption of non-domesticated animal species.

A research project on the impact of improving village chicken production on human diets and nutrition was carried out between April 2015 and June 2017 in response to the low frequency of consumption of animal-source food (84) and high

levels of indigenous chicken ownership in Timor-Leste (85). This research project employed mixed methods through a gender-sensitive lens to collect qualitative data from three rural villages in the eastern, central, and western regions of the country. These villages were involved in a pilot ND vaccination program for village chickens between November 2014 and January 2017, with quantitative data being collected from the three pilot villages and three matched control village not vaccinating against ND. The research project was conducted through The University of Sydney in collaboration with the Timor-Leste Ministries of Agriculture and Fisheries and Health and funded through The University of Sydney and the Australian Government, while the ND vaccination program was a collaboration between the Australian Department of Foreign Affairs and Trade and the Timor-Leste Ministry of Agriculture and Fisheries, and was funded by the Australian Government.

Participating households (56–71 households per village) were selected on the basis of having one child under the age of 2 years at the time of enrolment and were followed longitudinally for just over 2 years. Quantitative data on chicken flock management were collected monthly in pilot villages, as well as dietary diversity data and anthropometric measurements from mothers and children seasonally from all six villages. This study involved participation by consultation, where the external researcher identified problems and potential solutions though gathering qualitative data on household food availability, infant, and young child (IYC) feeding and chicken flock response to ND vaccination through annual KIIs and FGDs. Key informants included village and sub-village heads, cultural leaders, local and municipal health, and agricultural staff. Focus group discussions involved both young and old members of the community and were sex-disaggregated.

While Timor-Leste is typically described as only having two seasons (a wet and a dry season), the seasonal calendars created through KIIs and FGDs identified three agriculturally important seasons: the dry season; wet season; and less-wet season. This finding informed the timing of collection of quantitative data for this study and allowed a more nuanced study of seasonal impact on diets and animal-source food (ASF) consumption. Through the quantitative study, adult dietary diversity was found to be significantly lower in the dry and wet seasons compared to the less-wet season. The qualitative study complemented the quantitative study by exploring the reasons behind the differences in food consumption through the year, and revealed both seasonal and non-seasonal drivers for household animal-source food consumption (7). In these rural areas of Timor-Leste, most animal-source food consumption was reported to occur during social events. Non-seasonal events included marriages, illnesses or deaths, and events occurring at fixed times, such as national holidays. Seasonal events included the consumption of ASF when guests visit, typically during the dry season, and ritualistic offerings for maize planting and harvest occurring at the start and finish of the rainy season. Local chicken is the most frequently consumed animal-source food, due to their significance in sociocultural practices, as well as their availability and the preference for the taste and texture of local chicken meat. Other animal-source food consumption practices also follow a seasonal pattern due to changing environmental conditions. Some farmers consume more eggs during the dry season, when decreased foliage increases chick predation. Where allowed, the hunting of non-domesticated animals was found to be a common practice amongst men and boys, and occurred more frequently during the dry season.

Triangulation between qualitative and quantitative findings identified an animal-source food consumption practice that has public health implications due to food safety. Livestock are valuable household assets: in many LMICs, the slaughter and consumption of livestock during disease epidemics is a common way for households to mitigate losses. In this study, quantitative data showed that household chicken consumption increased by 10–35% (n = 30–77) during ND outbreaks, and KIIs and FGDs confirmed that slaughter and consumption of sick birds, or consumption of recently dead birds in good condition were common practices. For zoonotic poultry diseases that result in higher morbidity and mortality rates in humans, such as HPAI, this practice could pose a significant public health risk. Programs that reduce the prevalence of fatal endemic diseases of poultry and livestock not only increase the numbers of healthy animals, but may also reduce consumption of sick animals and zoonotic disease transmission.

Quantitative analysis of IYC diets and qualitative exploration of IYC animal-source food feeding practices revealed that although eggs were considered culturally acceptable foods for IYC and parents preferentially gave eggs to IYC over adults, meat was considered texturally too tough for IYC to digest. This has implications for livestock interventions aiming to increase the availability of meat for household consumption, particularly if the improvement of child nutrition is an intervention target.

Finally, over the course of this 2-year longitudinal study, the researchers observed changes in the enthusiasm of female participants to engage during FGDs. In contrast to male participants, women were initially reluctant to voice their opinions or concerns. Repeated visits over a longer timeframe fostered familiarity and trust between the researchers and the participants, and women were able to speak more freely at subsequent FGDs and so achieved interactive participation (86).

Key findings from this study were reported back to stakeholders, with separate meetings conducted with the Australian Government, the Timor-Leste Ministries of Agriculture and Fisheries and Health at national and regional levels, and with village leaders and participants. Findings were also presented to a wider group of stakeholders in Dili, including researchers, multilateral organizations, and local and international NGOs.

Uganda: Adapting PE to Understand Human Malnutrition

The Republic of Uganda is a landlocked country in East-Central Africa (87). It is bordered to the east by Kenya, to the north by South Sudan, to the west by the Democratic Republic of the Congo, to the south-west by Rwanda, and to the south by

Tanzania. The southern part of the country includes a substantial portion of Lake Victoria, shared with Kenya and Tanzania. Uganda is in the African Great Lakes region. Uganda also lies within the Nile basin, and has a varied but generally a modified equatorial climate. In 2013, over a third of young children were stunted, 6% wasted, 14% underweight, 49% anemic, and 38% were deficient in Vitamin A (88).

As presented above, the early development of PE occurred largely in remote pastoralist areas of east Africa, and a recent adaptation of experiences among veterinarians in the 1990s and 2000s was to use PE methods to improve understanding of acute malnutrition in children and mothers in Karamoja, Uganda (89). In common with many other pastoralist areas, Karamoja has long been characterized by unacceptably high levels of acute malnutrition in children, despite significant investment in human nutrition and food security programs in these areas over many years. In 2016, there were 24 "information-giving" nutrition projects or programs in Karamoja implemented by 17 organizations, but the level of global acute malnutrition was increasing (90). An initial, informal review of programming approaches and types of nutrition intervention in Karamoja indicated three possible weaknesses. First, implementing agencies seemed not to consider the marked seasonality in livelihoods and food availability in Karamoja; conventional nutrition surveys were conducted twice a year and provided point prevalence estimates for global acute malnutrition (GAM), but provided limited information on monthly or seasonal variation in GAM. Second, Karamoja was experiencing important changes in livelihoods, with many households with low livestock ownership. Third, the knowledge and experience of women in project design had been overlooked, and there was limited understanding of women's perceptions of the main causes of acute malnutrition or their preferences for nutrition interventions.

With this context in mind, in 2018 an analysis employing consultative participatory tools was designed that aimed to describe the seasonality of acute malnutrition in Karamoja, and women's knowledge on the cause of acute malnutrition. The study was funded by USAID, UK Aid, and Irish Aid, implemented in collaboration with the Karamoja Resilience Support Unit and had two main phases. There was an initial ethnographic phase to document how women in Karamoja described malnutrition and related factors in their own language. Then, drawing on the initial phase, two PE methods were designed. First, a monthly calendar method enabled women to illustrate monthly variations in rainfall, availability of main food types, workload, human births, human diseases, and acute malnutrition. This method was designed to compare monthly changes in these variables, and women were provided with 100 counters for each variable and asked to distribute the counters by month. Therefore, the method showed monthly patterns of each variable using a standard, but arbitrary scale, and did not aim to produce absolute measures. Second, a causal diagram that involved scoring of the main causes of acute malnutrition and illustrating any important relationships between these causes.

Among the key findings from this work was a hidden peak in acute malnutrition in January and February, which coincided with very limited availability of animal milk or availability of home-produced cereals. Nutritional status improved with the onset of rain, pasture growth, and resumption of milk production by livestock herds. This improved nutritional status was maintained and was supported by crop harvests toward the end of rainy months (**Figure 2**). As nutritional surveys were usually conducted in November or December, and then June or July, the surveys did not capture the peak in acute malnutrition in January and February. Women provided credible accounts of the causes of malnutrition (**Figure 3**). They explained that they thought malnutrition had two root causes: (i) the limited availability of livestock and milk; and (ii) social norms that overburden women with childcare responsibilities and finding and preparing food for the family. The women felt that these two root causes were interlinked and led to other issues and problems. Significantly, limited livestock ownership had a direct impact on food availability due to insufficient milk supply, but in a social context in which women were responsible for feeding the family. As households were forced to find more non-livestock sources of food and income, most of this burden fell to women. The non-livestock activities included crop production (frequently on small plots and with a high risk of rain failure) and a range of other activities that involved substantial effort for meager reward, and which hampered childcare. While women worked, their unweaned children remained at home under the care of siblings or other household members, with inadequate or no milk available to nourish them. Additional issues were linked to livestock-gender root causes, for example the loss of cattle affected men by negatively impacting on their self-identity and sense of purpose, and enabled them to spend more time in villages than previously with increased consumption of local brew and hard liquor. From the women's perspectives, this increased the risk of violence against them, and the likelihood unplanned pregnancies.

This example of PE methods showed how PE could be used to describe and explain multiple and complex food production and social factors that cause acute malnutrition, and which were difficult to capture using conventional nutritional surveys.

Vietnam: Application of PE to Avian Influenza and Antimicrobial Resistance Control

Vietnam in southeast Asia is a mountainous country bounded by China to the North and Laos and Cambodia to the West (91). With a population of more than 95.5 million, Vietnam has seen rapid economic growth since major economic and political reforms in 1986, transitioning it from a low income to a rising lower middle-income economy. The poverty rate is now below 6% (92). An agricultural policy of decollectivization commencing in 1988 allowed rural households to take long-term contracts on land, and rent or buy capital stock and working capital. This policy shift away from cooperatives has been credited for the return to family farming in Vietnam (93). Nearly 40% of land in Vietnam is dedicated to agricultural production and 43% of the population are engaged in agricultural activities (94).

Livestock-keeping in Vietnam is characterized by smallholdings; 89% of farms are small family farms and on

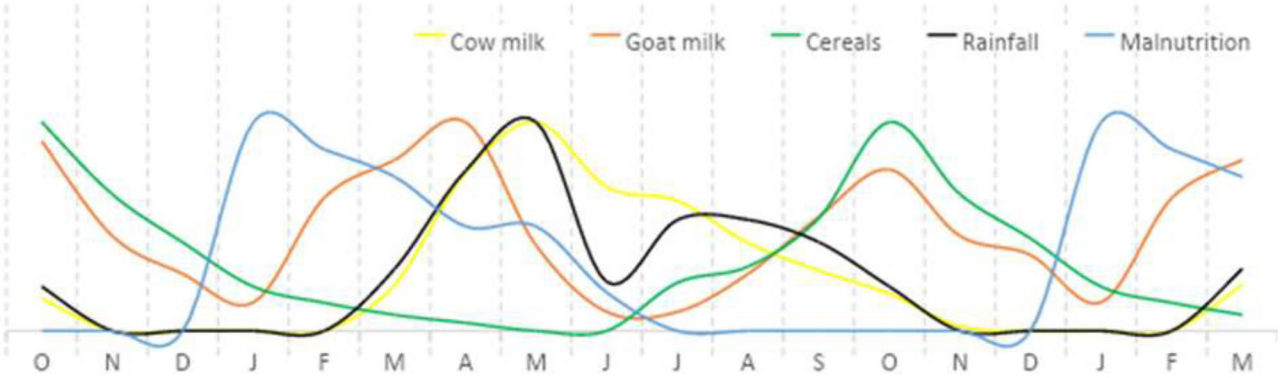


FIGURE 2 | Example of a monthly calendar for a "typical year" for child malnutrition, rainfall, and availability of livestock milk and cereals for human consumption, Karamoja, Uganda, 2018. Monthly calendars with 16 women's groups; in each group each variable was illustrated by distributing 100 counters across 12 months; summated scores from all 16 groups were used to construct the diagram, and the y-axis scale is arbitrary; the monthly calendar method was based on the Gregorian, solar calendar with 12 months; this example was constructed over an 18-month period to enable comparison of trends over consecutive end of year periods.

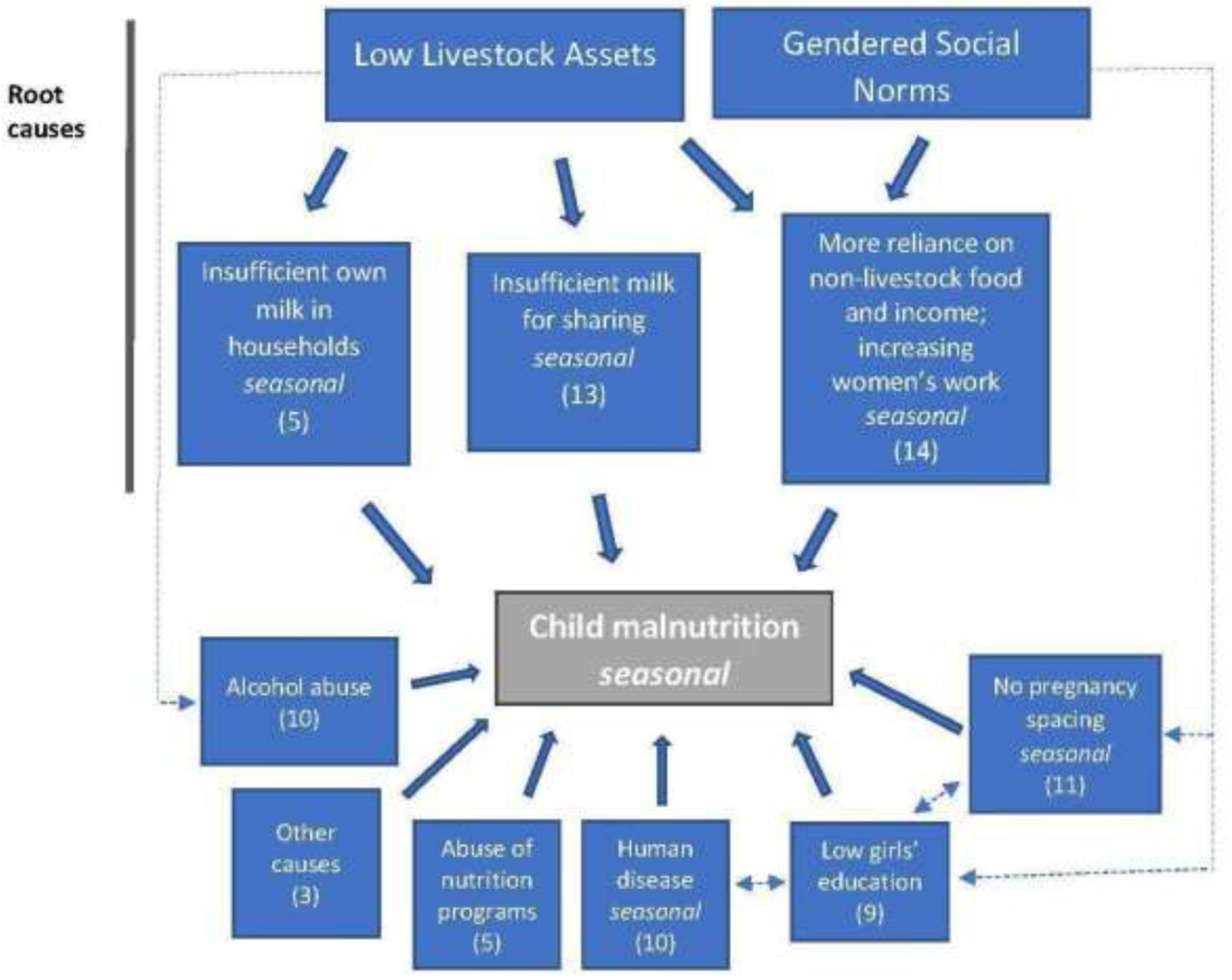


FIGURE 3 | Example of a causal diagram for child malnutrition, Karamoja, Uganda, 2018. The numbers in the boxes are the median scores for the relative importance of the causal factors.

average, pastoralists own 1.7 tropical livestock units (94). Livestock are often secondary sources of income after rice and other crops but nonetheless, form an important part of agricultural livelihoods. Overall, livestock account for around 5.9% of Vietnam's gross domestic product (95). According to the World Organization for Animal Health (OIE) Performance of Veterinary Services (PVS) report (96) the veterinary services in Vietnam continue to face many and complex challenges spanning governance (chain of command), training of veterinarians and paraveterinarians, and physical, financial and human resources. The OIE PVS report, itself developed in a participatory manner with the veterinary services, describes the need for improvements in stakeholder engagement as one of three cross-cutting priorities for improvement of the veterinary service. Participatory epidemiology offers a low-input approach to putting primary stakeholders, livestock-keepers at the center of animal disease research and development (2). This section outlines how PE was used to tackle the challenges of HPAI, and more recently, antimicrobial resistance (AMR) in Vietnam, and describes how PE displayed particular utility in better understanding formal and informal communication channels between farmers and other animal health stakeholders, enabling effective design of ensuing research and interventions.

In 2003, Vietnam saw the beginning of a devastating HPAI epidemic. In a country dominated by smallholder poultry systems, surveillance was a daunting task. In order to understand the ways information about suspect HPAI cases flowed, between 2012 and 2013, CIRAD-The French Agricultural Research Center for International Development, the Hanoi University of Agriculture, the Nong Lam University and the National Institute of Veterinary Research of Vietnam engaged multiple stakeholder groups in a consultative PE process. Focus Group Discussions including semi-structured group interview and PP with poultry farmers were used as a starting point. Further participants were identified by snowball; farmers were asked who they communicated with when they suspected HPAI and when they mentioned a new participant group, the research team asked for any particular names. The team then contacted these people for interview. Proportional piling was used to quantify the relative likelihood of sharing information with each participant group mentioned. Groups identified included both people in public roles (government veterinarians) and private roles (feed and chick sellers, veterinary medicine sellers, veterinary technicians of feed companies, and pharmaceutical companies). Importantly, it was found that people in private roles had greater access to information in the face of suspect HPAI outbreaks compared with the government surveillance system, which "appeared as peripheral in the information sharing network" despite mandatory reporting. In fact, the local private workers were largely responsible for spreading the information to distant areas, acting as somewhat of an early warning system to farmers. Using this snowball technique to follow the flow of information, it became apparent that to enhance passive surveillance of HPAI there was a need for greater communication links between private and public veterinary services (97). Building on this study, further PE approaches were used to document the perceived benefits and costs of a passive surveillance system for HPAI. The authors explained, PE was useful for integrating economic and non-economic costs and benefits as well as stakeholder perceptions. Farmers were found to face uncertainty in transaction and outcome costs associated with notifying the government of suspicious cases. In this PE process, while the researchers defined the research problem, the truly consultative nature of their approach was evidenced by how they listened to stakeholders, basing their recommendations on the stakeholders' responses. A key recommendation to the government was consistency in response to notification, such as rules for compensation. One of the benefits of engaging multiple stakeholders in this approach was that it highlighted some agreement in perceived costs to reporting; all stakeholders (farmers, veterinary authorities, and private, upstream participant groups) anticipated a drop in market prices if knowledge of HPAI suspicions were released. The findings suggested that the benefits for all stakeholders to report disease outweighed the benefits of silence only if the market for selling diseased animals did not exist. The recommendation arising from this finding was that the poultry value chain needed greater quality control (98). The case of HPAI in Vietnam demonstrates the harmony between PE's "ground up" approach and the interrogation and augmentation of passive surveillance. As in South Sudan, PE enhanced the under-resourced government surveillance system.

In recent years, many stakeholders involved in the HPAI response in Vietnam have been involved in responding to other One Health challenges (99) including important emerging infectious diseases caused by antibiotic-resistant bacteria. Antimicrobial resistance, the ability of microbes to evade antimicrobials and therefore render them ineffective, is a natural phenomenon but is rapidly increasing due to the overuse and misuse of antimicrobials in humans and animals. Veterinarians, animal caretakers, doctors, and their patients are called to be better stewards of antimicrobials, to slow the increase in AMR. Described as one of the greatest health threats of our time (100), AMR is considered a priority challenge for human and animal health sectors in Vietnam (101). Overuse and misuse of antimicrobials does not occur in a vacuum. Especially in LMICs such as Vietnam where antimicrobials can often be purchased with no prescription, an understanding of the social and socio-economic context is crucial to designing better policy and implementing change (102). To this end, two initiatives, one in Southern Vietnam, and one in Northern Vietnam have applied participatory approaches and methods to the challenge of veterinary AMR. Both approaches engaged multiple stakeholders and explored antimicrobial use from economic and non-economic angles.

In Northern Vietnam, from 2016 to 2018, PE methods were used in a sequential mixed methods design to understand and look for ways to improve veterinary antimicrobial stewardship in family farming. The study was led by The University of Queensland and the International Livestock Research Institute in collaboration with the Hanoi University of Public Health and Thai Nguyen University of Agriculture and Fisheries. The research was funded by the CGIAR Research Program on Agriculture for Nutrition and Health and The Australian Government's Research Training Program. As in the HPAI study

above, farmer FGDs were used as a first step, to identify further relevant participant groups and identify themes for further study. Farmers were asked a broad question, "What happens when a pig gets sick on your farm?" As the farmers mentioned the steps taken by themselves and other people they interacted with, they were written on cards to prompt discussion. In contrast to the study in southern Vietnam below, consensus was not sought. Rather, the group was probed to elicit the greatest diversity of responses to sick pigs in different scenarios until no new responses were provided (saturation). The reasons for various steps taken, for example, financial or physical constraints, beliefs and experience, were taken forward as themes for further study. Participant groups identified in the FGDs were also interviewed using semi-structured interview. The main groups identified were government veterinarians and private community animal healthcare workers. The findings were used to develop semi-quantitative survey tools for farmers and the additional participant groups identified. After the implementation of the survey and preliminary data analysis, the findings were brought back to the community for interpretation and development of a list of proposed interventions. This final stage included PE activities with individual participant groups, followed by a combined workshop of farmers, private, and public animal healthcare workers. Through this process, points of convergence and divergence were explored, and an agreed list of proposed interventions to improve antimicrobial stewardship finalized. As in the HPAI study, there were many points of common understanding and agreement. The community then presented these agreed recommendations to local and regional government and other external stakeholders. Using this adaptive, multi-stage process, the engagement of community groups moved from consultative participation in the first FGDs to participation in information giving in the surveys, and finally toward interactive and functional participation in joint analysis and proposal of interventions to external stakeholders including local authorities. The process was still dependent on external facilitators. Major decisions regarding governance of antimicrobials were proposed to be made by those people in positions of power, external to the community. However, some local decisions and plans, such as those to improve animal husbandry, were made by the participants. Through maintaining farmers as central stakeholders in the research and including other groups the farmers identified as important, this participatory process highlighted opportunities to improve antimicrobial stewardship that were agreeable to all (103).

More recently, from December 2017 to March 2018, during a Wellcome Trust-funded study in the southern Mekong Delta region of Vietnam, poultry farmers, veterinary drug shop owners, government veterinarians, and animal healthcare workers were engaged in a two-stage, mixed consultative PE and Q-sorting process. The first stage was "collective interview," considered by the authors as a more appropriate term than focus group because the groups were heterogenous and they were seeking consensus rather than exploring alternative points of view. Before consensus was sought, however, PE methods were employed to allow participants to freely explore topics. Interview guides were semi-structured and PE methods included pairwise ranking, timelines, PP, and flow-charts to characterize poultry diseases, their prevention and control, identify sources of advice, and determine the timing and positive and negative opinions around antimicrobial use. These themes were chosen by the researchers based on knowledge of antimicrobial use and AMR. These data were used to develop a series of statements for use with a Q-sorting tool. This tool involved individual participants indicating on a scale how strongly they agreed or disagreed with each statement. The interviews allowed a more nuanced understanding of diversity of opinions and also provided opportunities for triangulation to verify findings, an important aspect of PE (104).

DISCUSSION

Within the veterinary arena, participatory epidemiology emerged in the final two decades of the twentieth Century, initially in association with small-scale, community-based development projects and subsequently playing a key role in the global effort to eradicate rinderpest (4). Working with communities in areas where animal health services were weak or non-existent was vital to understanding disease dynamics and opportunities to implement cost-efficient control programs (1). Over two decades later, resource shortages and competing priorities mean that vaccine-preventable animal diseases continue to kill huge—and in many cases undocumented—numbers of animals across the globe (105, 106), antimicrobial resistance has grown significantly (107) and food insecurity is rising (108).

Participatory epidemiology capitalizes on what is known and encourages communities to use their own knowledge of and skills with the animals they keep, the infectious diseases affecting their animals and the human diseases which can be acquired from their animals and vice versa. Indigenous knowledge which emerges from the experience of keeping the animals over long time periods enables animal keepers to define the clinical signs, salient lesions and epidemiological behavior in their own words which frequently have parallel meaning with technically employed terms (109). A failure to incorporate this local knowledge and experience may result in wrong conclusions and interventions which can fail to effectively and sustainably address the problem. Therefore, participatory epidemiological research provides a more comprehensive and diverse knowledge relevant for catalyzing positive change in the community toward solving their own problem in sustainable manner (25). The type of the approach and methods used in participatory epidemiology, when correctly employed, ensure inclusion in terms of gender, and different levels of education of the participants in seeking solutions to community problems. This enriches the information gathered (including more convention epidemiological data) and makes it specific to that locality, which are both important aspects in designing appropriate problem-solving strategies.

Participatory approaches provide an opportunity for all involved to agree on objectives. For example, in both Pakistan and Indonesia, endemic diseases such as haemorrhagic septicaemic and rabies were key priorities for communities, while in Timor-Leste, control of endemic ND may assist in decreasing

the frequency of risky consumption practices. Openness by project funders and implementers to identify opportunities to control diseases of both local and global importance can help to build trust between participants and lead to more sustainable disease prevention, surveillance and control. These findings are in line with recommendations by Allepuz et al. (3) that PE techniques be employed to enhance dialogue between producers and national veterinary services.

Participatory research for collaborative, just action is useful for establishing unity in research and development goals, indicators for monitoring, and for understanding why interventions may or may not have been successful, i.e., it can be integral to learning. The PE research on AMR conducted in Northern Vietnam is an example of where authorities were alerted to on-the-ground challenges and opportunities in controlling an important emerging infectious disease. Participatory methodologies are a crucial component of all research, interventions or programmes that are likely to be affected by multiple factors. The inclusion of gender-sensitive approaches during the application of PE techniques increases the likelihood that the perspectives of more marginalized households and more vulnerable household members are heard and acted upon (2, 23).

The effectiveness of the One Health approach to infectious disease control has been greatly enhanced by incorporating social scientists and relevant participatory activities involving multiple sectors as outlined in the case studies from Indonesia, South Sudan, Tanzania, and Vietnam. The US Centers for Disease Control and Prevention define One Health as a collaborative, multisectoral, and transdisciplinary approach—working at the local, regional, national, and global levels—with the goal of achieving optimal health outcomes recognizing the interconnection between people, animals, plants, and their shared environment (110). The recognition of the importance of the social elements in this definition has been a major advancement in One Health implementation. In addition to social and political scientists, economists can play a role in ensuring that key findings, relating to both technical and policy aspects, are effectively presented to and addressed by senior decision-makers as presented in the Indonesian case study.

The combination of growing pressures on planetary (111) and social (112) boundaries and desperately inadequate funding for agricultural (especially livestock and aquaculture) research and development (113)—now further impacted by the contraction of the global economy due to COVID-19—makes it even more important that all resources are utilized as efficiently as possible. Projects and programs that run over a longer timeline can learn from and respond to community knowledge and priorities and provide increased opportunities for findings to be incorporated into policy and policy implementation frameworks. Longer-running activities also have a greater chance of achieving truly interactive participation, leading to self-mobilization as envisaged by Pretty (16). The PE activities presented in the case studies from Indonesia, Pakistan, South Sudan, and Uganda largely employed participation by consultation. Engagement with specific communities was generally over a short period with the information extracted contributing to larger national goals but rarely discussed with participants subsequently. These PE activities tended to be associated with larger, time-bound projects designed without community engagement. In the Karamoja case study in Uganda, the general approach of human nutrition programs was top-down, with women not being consulted or listened to, but expected to adopt program messages and change their behaviors. The PE study was more participatory relative to other programs and the general development context, and the first time that women's views had been documented, which should be considered a positive step. By contrast, the PE activities presented in the Tanzanian, Timor-Leste, and Vietnam case studies ran over 2–5 years and involved ongoing engagement with the same communities. Information and analyses were presented back to communities for discussion and subsequent action. Where veterinary research initiatives are concerned, funds are rarely allocated to monitoring and evaluating collective action, so, even if some of the PE research leads to self-mobilization, it has rarely been measured and reported. Consequently, even with the best of intentions, PE research is often constrained to participation by consultation, due to external forces (e.g., funder priorities and timelines). In terms of supporting functional community mobilization, the findings from the case studies suggest that project longevity may be more important than the size of the budget. Shorter term, large budget projects, may generate information but such projects rarely lead to transformation at the community level. Designing projects that run over longer periods, incorporate collective and reflexive learning through continuous evaluation (114) and are adaptive in line with findings, are more likely to achieve functional participation that can lead to self-mobilization (115, 116).

The triangulation of information using a combination of synergistic qualitative and quantitative tools provides an excellent opportunity to assess the robustness of the data collected and frequently provides insights into why the findings arose. This was amply demonstrated in the case studies from Tanzania, Timor-Leste, and Uganda in relation to household nutrition security, a complex and challenging issue that is advanced through effective participation and collaboration between communities, government agencies, and research and development personnel. Similar approaches are being refined in relation to water, sanitation, and hygiene research and development activities (117). In relation to the use of PE in infectious disease and EID outbreaks, including laboratory diagnosis as a component of the suite of triangulation tools employed is important (36). This was amply demonstrated in the South Sudan case study where laboratory testing confirmed the presence of HPAI H5N1 and not Newcastle disease as had been suspected by local producers.

As summarized above, a significant number of educational and training materials on PE techniques are freely available online and mostly in English. The need for additional material has been recognized, as has increasing opportunities for discussions relating to the inclusion of PE activities into regular national veterinary services programs (3). The more robust use of gender-sensitive methodologies and a gender lens (such as routine gender-disaggregation of data collection and analysis, the application of same gender discussion groups, gender-sensitive training curricula and methodologies, and empowerment tools)

would enable participatory approaches to better address issues of socio-economic-, gender- and language-based differences.

CONCLUSIONS AND RECOMMENDATIONS FOR THE WAY FORWARD

Development literature, and increasingly epidemiological literature, abounds with references to the importance and effectiveness of participatory approaches. Despite this, participatory approaches remain less commonly utilized and inadequately resourced compared to more top-down approaches which all too-frequently fail to be successful in the long-term. In many parts of the world, vaccine-preventable animal diseases remain uncontrolled, contributing to food and nutrition insecurity, foodborne disease and antimicrobial resistance. More determined adherence to the fundamental principles of participatory epidemiology that communities must be actively involved "in defining and prioritizing problems, and in the development of solutions to them," as defined by Catley et al. (2), is vital. This will require adaptive management of projects that are of a sufficient duration for trust and effective collaboration to develop between partners. The incorporation of gender-sensitive participatory impact assessments into activities will assist with measuring the degree to which objectives have been met and key outcomes achieved for all stakeholders.

AUTHOR CONTRIBUTIONS

RA conceived the paper, took the lead with writing the general sections, and contributed to all sections. SA wrote the case study on Pakistan. AA wrote the study on South Sudan. BB contributed to the Tanzanian section study. TC wrote the section on Vietnam, AG and MH wrote the study on Indonesia. ER contributed to the Tanzanian study. JW wrote the Timor-Leste study. AC wrote the Ugandan study. All authors contributed to the revision process for all sections.

FUNDING

Support for activities is gratefully acknowledged for the work in: Indonesia funded by the Government of Indonesia and USAID (FAO Project No. OSRO/INS/501/USA); Pakistan funded by the Government of Pakistan, FAO, and the European Union; Tanzania by the Governments of Tanzania and Australia; Timor-Leste by the Government of Timor-Leste, The University of Sydney and the Government of Australia; Uganda by the Government of Uganda, Irish Aid, UK Aid, UNICEF, USAID, and WFP; and the Northern Vietnam AMR study, by the CGIAR Research Program on Agriculture for Nutrition and Health. TC wishes to acknowledge support from the Australian Government Research Training Program Scholarship and the Crawford in Queensland Postgraduate Student Award.

REFERENCES

1. Catley A, Mariner J. *Where There is No Data: Participatory Approaches to Veterinary Epidemiology in Pastoral Areas of the Horn of Africa*. International Institute for Environment and Development Issue Paper No. 110. (2002). Available online at: https://pubs.iied.org/pdfs/9179IIED.pdf (accessed December 28, 2019).
2. Catley A, Alders RG, Wood JL. Participatory epidemiology: approaches, methods, experiences. *Vet J.* (2012) 191:151–60. doi: 10.1016/j.tvjl.2011.03.010
3. Allepuz A, de Balogh K, Aguanno R, Heilmann M, Beltran-Alcrudo D. Review of participatory epidemiology practices in animal health (1980–2015) and future practice directions. *PLoS ONE*. (2017) 12:e0169198. doi: 10.1371/journal.pone.0169198
4. Catley A. Participatory epidemiology: reviewing experiences with contexts and actions. *Prev Vet Med*. (2020) 180:105026. doi: 10.1016/j.prevetmed.2020.105026
5. de Bruyn J, Bagnol B, Darnton-Hill I, Maulaga W, Thomson P, Alders R. Characterising infant and young child feeding practices and the consumption of poultry products in rural Tanzania: a mixed methods approach. *Matern Child Nutr*. (2017) 14:e12550. doi: 10.1111/mcn.12550
6. de Bruyn J, Msuya J, Ferguson E. Evaluating pictorial charts as a means of collecting participant-recorded data on household dietary diversity in low-literacy communities in Tanzania. *Br J Nutr*. (2019) 122:1432–40. doi: 10.1017/S0007114519002587
7. Wong JT, Bagnol B, Grieve H, Jong J, Li M, Alders RG. Factors influencing animal-source food consumption in Timor-Leste. *Food Secur*. (2018) 10:740–62. doi: 10.1007/s12571-018-0804-5
8. Suseno PP, Jatikusumah A, Husein WF, Sawitri E, Wicaksono A, Gozali A, et al. Global health in action: One Health approach to rabies surveillance and control in Indonesia. In: *Global Health Security Conference*. Sydney, NSW (2019).
9. Ryser-Degiorgis M-P. Wildlife health investigations: needs, challenges and recommendations. *BMC Vet Res*. (2013) 9:223. doi: 10.1186/1746-6148-9-223
10. Bagnol B. Improving village cicken production by employing effective gender sensitive methodologies. In: Alders RG, Spradbrow PB, Young MP, editors. *Village Chickens, Poverty Alleviation and Sustainable Control of Newcastle Disease*. Canberra, NSW: Australian Center for International Agricultural Research. ACIAR Proceedings No. 131 (2009). p. 35–42. Available online at: http://www.aciar.gov.au/publication/PR131 (access January 30, 2020).
11. Bagnol B. Advocate gender issues: a sustainable way to control newcastle disease in village chickens. *Good Practices for Family Poultry Production*. International Network for Family Poultry Development; IFAD; FAO. GPFPP Note No. 3. (2013). Available online at: http://www.fao.org/docrep/018/aq630e/aq630e.pdf (accessed January 30, 2020).
12. Alders RG, Bagnol B. Effective communication: the key to efficient HPAI prevention and control. *Worlds Poult Sci J*. (2007) 63:139–47. doi: 10.1017/S0043933907001365
13. Bagnol B, Naysmith S, de Bruyn J, Wong J, Alders R. Effective animal health programming requires consideration of and communication with those at the human–animal interface. *CAB Rev*. (2016) 11:1–7. doi: 10.1079/PAVSNNR201611030
14. Bagnol B. Conducting participatory monitoring and evaluation. In: *Decision Tools for Family Poultry Development*. FAO Animal Production and Health Guidelines No. 16. Rome: FAO (2014). Available online at: http://www.fao.org/docrep/019/i3542e/i3542e.pdf (accessed January 30, 2020).
15. Freire P. *Pedagogy of the Oppressed*. New York, NY: Herder and Herder (1970).
16. Pretty J. Alternative systems of inquiry for a sustainable agriculture. *IDS Bull*. (1994) 25:37–48. doi: 10.1111/j.1759-5436.1994.mp25002004.x
17. Carr A, Wilkinson R. Beyond participation: boundary organizations as a new space for farmers and scientists to interact. *Soc Nat Resour*. (2005) 18:255–65. doi: 10.1080/08941920590908123

18. Chambers R. *Rural Development: Putting the Last First*. London: Longmans (1983).
19. Scarborough V, Killough S, Johnson DA, Farrington J. *Farmer-Led Extension: Concepts and Practices*. London: Intermediate Technology Publications (1997)
20. Garforth C, Angell B, Archer J, Green K. *Improving Farmers' Access to Advice on Land Management: Lessons From Case Studies in Developed Countries*. London: Agricultural Research and Extension Network; Overseas Development Institute (2003).
21. Alex G, Rivera W, editors. *Demand-Driven Approaches to Agriculture Extension: Case Studies of International Initiatives*. Agriculture and Rural Development Discussion Paper No. 10. Extension Reform for Rural Development. Washington, DC: World Bank (2004). Available online at: http://documents.worldbank.org/curated/en/184731468298459108/Demand-driven-approaches-to-agriculture-extension-Case-studies-of-international-initiatives-vol-3 (accessed June 15, 2020).
22. Howell KE. *An Introduction to the Philosophy of Methodology*. London: SAGE Publications Ltd. (2013).
23. Bagnol B, Alders R, McConchie R. Gender issues in human, animal and plant health using an ecohealth perspective. *Environ Nat Resour Res.* (2015) 5:62–76. doi: 10.5539/enrr.v5n1p62
24. Blouin-Genest G, Miller A. The politics of participatory epidemiology: technologies, social media and influenza surveillance in the US. *Health Policy Technol.* (2017) 6:192–7. doi: 10.1016/j.hlpt.2017.02.001
25. Bach M, Jordan S, Hartung S, Santos-Hövener C, Wright MT. Participatory epidemiology: the contribution of participatory research to epidemiology. *Emerg Themes Epidemiol.* (2017) 14:2. doi: 10.1186/s12982-017-0056-4
26. Freudenberger KS. *Rapid Rural Appraisal and Participatory Rural Appraisal: A Manual for CRS Field Workers and Partners*. MD: Catholic Relief Services (2011). Available online at: https://www.crs.org/sites/default/files/tools-research/rapid-rural-appraisal-and-participatory-rural-appraisal.pdf (accessed April 18, 2020).
27. Chambers R. The origins and practice of participatory rural appraisal. *World Dev.* (1994) 22:953–69. doi: 10.1016/0305-750X(94)90141-4
28. Maalim AD. Participatory rural appraisal techniques in disenfranchised communities: a Kenyan case study. *Int Nurs Rev.* (2006) 53:178–88. doi: 10.1111/j.1466-7657.2006.00489.x
29. Cornwall A, Pratt G. The use and abuse of participatory rural appraisal: reflections from practice. *Agric Human Values.* (2011) 28:263–72. doi: 10.1007/s10460-010-9262-1
30. Solomon BD. Environmental reviews and case studies: socioeconomic analysis options for pesticides management in developing countries: a review. *Environ Pract.* (2015) 17:57–68. doi: 10.1017/S1466046614000507
31. Baum F, MacDougall C, Smith D. Participatory action research. *J Epidemiol Community Health.* (2006) 60:854–7. doi: 10.1136/jech.2004.028662
32. Schneider B. Participatory action research, mental health service user research, and the hearing (our) voices projects. *Int J Qual Methods.* (2012) 11:152–65. doi: 10.1177/160940691201100203
33. Ellsberg M, Heise L. *Researching Violence Against Women: A Practical Guide for Researchers and Activists*. Washington, DC: World Health Organization (2005).
34. Mwongera C, Shikuku KM, Twyman J, Laderach P, Ampaire E, van Asten P, et al. Climate smart agriculture rapid appraisal (CSA-RA): a tool for prioritizing context-specific climate smart agriculture technologies. *Agric Syst.* (2017) 151:192–203. doi: 10.1016/j.agsy.2016.05.009
35. Mariner JC, Paskin R. *Manual on Participatory Epidemiology—Method for the Collection of Action-Oriented Epidemiological Intelligence*. Rome: Food and Agriculture Organization of the United Nations (2000). Available online at: http://www.fao.org/3/x8833e/x8833e00.htm (accessed November 30, 2019).
36. Catley A. *Participatory Epidemiology: A Guide for Trainers*. Nairobi: African Union/InterAfrican Bureau for Animal Resources (2005). Available online at: https://www.livestock-emergency.net/userfiles/file/veterinary-services/Catley-2005.pdf (accessed November 28, 2019).
37. Catley A, Burns J, Abebe D, Suji O. *Participatory Impact Assessment: A Guide for Practitioners*. Massachusetts: Feinstein International Center Tufts University (2008). Available online at: https://www.researchgate.net/publication/257816175_Participatory_Impact_Assessment_A_Guide_for_Practitioners (accessed January 29, 2020).
38. Ahlers C, Alders RG, Bagnol B, Cambaza AB, Harun M, Mgomezulu R, et al. *Improving Village Chicken Production: A Manual for Field Workers and Trainers*. ACIAR Monograph No. 130. Canberra, NSW: Australian Center for International Agricultural Research (2009). Available online at: https://www.aciar.gov.au/node/9741 (accessed December 27, 2019).
39. Dunkle S, Mariner J. *Participatory Epidemiology: A Toolkit for Trainers*. Nairobi; Kenya: ILRI (2013). Available online at: https://cgspace.cgiar.org/handle/10568/35216 (accessed December 30, 2019).
40. Wikipedia. *Pakistan*. Wikipedia (2020). Available online at: https://en.wikipedia.org/wiki/Pakistan (accessed April 28, 2020).
41. Afzal M. Re-designing smallholder dairy production in Pakistan. *Pak Vet J.* (2010) 30:187–90.
42. Ahmad T, Ishaq K, Fiaz M, Parveen R, Safdar A. Small ruminant production systems: status and prospectus for improvement in Potohar Region of Pakistan. *Pak J Sci.* (2016) 68:253–8.
43. Kunbhar HK, Memon A, Bhutto A, Rajput ZI, Suthar V, Memon A, et al. Study on female reproductive performance of Kamohri goat managed under traditional management conditions in district Hyderabad, Sindh, Pakistan. *Int J Adv Res Biol Sci.* (2016) 3:251–60. Available online at: http://s-o-i.org/1.15/ijarbs-2016-3-3-32
44. Ali SN, Khan SA, Mohammed A, Kazmi SU, Mariner J. The association of disease with animal production systems and its socio-economic impact in district Tharparkar Sindh Pakistan. *Int J Biol Biotechnol.* (2018) 15:383–93.
45. FAO/OIE. *Joint FAO/OIE Committee on Global Rinderpest Eradication. Final Report. May 2011*. Food and Agriculture Organisation of the United Nations; World Organisation for Animal Health (2011). Available online at: https://www.oie.int/doc/ged/D10943.PDF (accessed April 28, 2020).
46. Ali SN, Asif M, Rehman A, Jat LA, Ali Q, Hussain M. Participatory surveillance of livestock diseases in district Karachi–Pakistan. *Int J Agric Biol.* (2006) 8:652–6.
47. Mariner JC, Hendrickx S, Pfeiffer DU, Costard S, Knopf L, Okuthe S, et al. Integration of participatory approaches into surveillance systems. *Revue Scientifique et Technique-Office International des Epizooties.* (2011) 30:653–9. doi: 10.20506/rst.30.3.2065
48. FAO. *Pakistan's Victory Over Rinderpest*. Rome: Food and Agriculture Organization of the United Nations (2003). Available online at: http://www.fao.org/english/newsroom/news/2003/13841-en.html (accessed January 15, 2020).
49. Zahur AB, Irshad H, Hussain M, Ullah A, Jahangir M, Khan MQ, et al. The epidemiology of peste des petits ruminants in Pakistan. *Revue Scientifique et Technique-Office International des Epizooties*. (2008) 27:877–84. doi: 10.20506/rst.27.3.1848
50. Jones B, Ameri AA, Mogga NK, Letereuwa SP. *Rinderpest Eradication in Southern Sudan: The Final Stages From 2001 to 2007*. Technical Report. Juba: Ministry of Animal Resources and Fisheries and Vétérinaires sans Frontières (VSF) (2010). Available online at: https://www.researchgate.net/publication/289538827_Rinderpest_eradication_in_Southern_Sudan_the_final_stages_from_2001_to_2007 (accessed December 31, 2019).
51. AVMA. *One Health: A New Professional Imperative. One Health Initiative Task Force: Final Report*. American Veterinary Medical Association (2008). Available online at: https://www.avma.org/sites/default/files/resources/onehealth_final.pdf (accessed April 22, 2020).
52. Mariner J, Jones BA, Hendricks S, El Masry I, Jobre Y, Jost CC. Experiences in participatory surveillance and community-based reporting systems for H5N1 highly pathogenic avian influenza: a case study approach. *EcoHealth*. (2014) 11:22–35. doi: 10.1007/s10393-014-0916-0
53. WHO. WHO promotes one-health approach to enhance response to zoonoses in South Sudan. World Health Organization; Regional Office for Africa (2018). Available online at: https://www.afro.who.int/news/who-promotes-one-health-approach-enhance-response-zoonoses-south-sudan (accessed December 29, 2019).
54. Fasanmi OG, Odetokun IA, Balogun FA, Fasina FO. Public health concerns of highly pathogenic avian influenza H5N1 endemicity in Africa. *Vet World.* (2017) 10:1194–204. doi: 10.14202/vetworld.2017.1194-1204
55. GoSS. *Rift Valley Fever (RVF) Outbreak—Yirol East, Eastern Lakes State.* Republic of South Sudan: GoSS. Situation Report (Sitrep No. 6) as at 17.00 Hours (2018).

56. Barasa M, Catley A, Machuchu D, Laqua H, Puot E, Tap Kot D, et al. Food and mouth disease vaccination in South Sudan: benefit-cost analysis and livelihoods impact. *Transbound Emerg Dis.* (2008) 55:339–51. doi: 10.1111/j.1865-1682.2008.01042.x
57. Catley A. *Training Course in Participatory Epidemiology*. Ganyiel: Summary Report. Suisse: Vétérinaires sans Frontières (VSF) (2005). Available online at: http://www.vetwork.org.uk/userfiles/Catley-PE-S-Sudan%20summary-report.pdf (accessed January 29, 2020).
58. Wikipedia. *Indonesia*. Wikipedia (2020). Available online at: https://en.wikipedia.org/wiki/Indonesia (accessed April 22, 2020).
59. Salikha A. *Meet The 10 Megadiverse Countries in The World*. SEAsia (2018). Available online at: https://seasia.co/2018/01/29/meet-the-10-megadiverse-countries-in-the-world (accessed April 22, 2020).
60. Azhar M, Lubis AS, Siregar ES, Alders RG, Brum E, McGrane J, et al. Participatory disease surveillance and response in Indonesia: strengthening veterinary services and empowering communities to prevent and control highly pathogenic avian influenza. *Avian Dis.* (2010) 54:749–53. doi: 10.1637/8713-031809-Reg.1
61. Alders RG, Bagnol B, Brum E, Lubis AS, Young MP. Continuing education in the prevention and control of HPAI: a case study on Indonesia. *Worlds Poultry Sci J.* (2009) 65:529–31. doi: 10.1017/S0043933909000397
62. FAO. *FAO Emergency Center for Transboundary Animal Diseases. Annual report 2018—Healthy Animals Healthy People*. Jakarta (2019). Available online at: http://www.fao.org/3/ca5749en/CA5749EN.pdf (accessed March 20, 2020).
63. FAO. *The Economic Case to Scale-up the Zoonoses Prevention Control Programme Using the One Health Approach in Indonesia: Policy Brief. Strengthening National Capacity to Prevent and Control Emerging and Re-emerging Pandemic Threats Including Influenza A in Indonesia (OSRO/INS/501/USA)*. Jakarta: Food and Agriculture Organization of the United Nations (2019).
64. Pertanian K. *Masterplan Nasional Pemberantasan Rabies di Indonesia*. Jakarta: Kementerian Pertanian (2019).
65. Central Intelligence Agency. *CIA World Factbook*. CIA (2020). Available online at: https://www.cia.gov/library/publications/the-world-factbook/geos/tz.html (accessed July 4, 2020).
66. Sunguya BF, Zhu S, Mpembeni R, Huang J. Trends in prevalence and determinants of stunting in Tanzania: an analysis of Tanzania demographic health surveys (1991–2016). *Nutr J.* (2019) 18:85. doi: 10.1186/s12937-019-0505-8
67. Edwin P, Azage M. Geographical variations and factors associated with childhood diarrhea in Tanzania: a national population based survey 2015–16. *Ethiop J Health Sci.* (2019) 29:513–24. doi: 10.4314/ejhs.v29i4.13
68. Masanja H, Mongi P, Baraka J, Jackson B, Kisisiwe Y, Manji K, et al. Factors associated with the decline in under five diarrhea mortality in Tanzania from 1980–2015. *J Global Health.* (2019) 9:020806. doi: 10.7189/jogh.09.020806
69. Alders R, Aongola A, Bagnol B, de Bruyn J, Kimboka S, Kock R, et al. Using a One Health approach to promote food and nutrition security in Tanzania and Zambia. *Planet@Risk.* (2014) 2:187–90.
70. Bagnol B, Clarke E, Li M, Maulaga W, Lumbwe H, McConchie R, et al. Transdisciplinary project communication and knowledge sharing experiences in Tanzania and Zambia through a One Health lens. *Front Public Health*. (2016) 4:10. doi: 10.3389/fpubh.2016.00010
71. Francis SC, Lees SS, Andrew B, Zalwango F, Vandepitte J, Ao T, et al. Vaginal practices diary: development of a pictorial data collection tool for sensitive behavioral data. *Sex Transm Dis.* (2012) 39:614–21. doi: 10.1097/OLQ.0b013e3182515fe4
72. Francis SC, Baisley K, Lees SS, Andrew B, Zalwango F, Seeley J, et al. Vaginal practices among women at high risk of HIV infection in Uganda and Tanzania: recorded behaviour from a daily pictorial diary. *PLoS ONE.* (2013) 8:e59085. doi: 10.1371/journal.pone.0059085
73. EFSA. Scientific report on the European Union One Health 2018 zoonoses report. *EFSA J.* (2019) 17:276. doi: 10.2903/j.efsa.2019.5926
74. Oloruntoba EO, Folarin TB, Ayede AI. Hygiene and sanitation risk factors of diarhoeal disease among under-five children in Ibadan, Nigeria. *Afr Health Sci.* (2014) 14:1001–2. doi: 10.4314/ahs.v14i4.32
75. Kabhele S, New-Aaron M, Kibusi SM, Gesase AP. Prevalence and factors associated with diarhoea among children between 6 and 59 months of age in Mwanza City Tanzania. *J Trop Pediatr.* (2018) 64:523–30. doi: 10.1093/tropej/fmx109
76. Havelaar AH, Kirk MD, Torgerson PR, Gibb HJ, Hald T, Lake RJ, et al. World Health Organization global estimates and regional comparisons of the burden of foodborne disease in 2010. *PLOS Med.* (2015) 12:e1001923. doi: 10.1371/journal.pmed.1001923
77. Rukambile E, Sintchenko V, Muscatello G, Alders R. Food safety and animal-source food: emerging challenges in resource-limited settings. Sydney, NSW: Global Health Security Conference (2019).
78. Rukambile E, Muscatello G, Maulaga W, Bagnol B, Stellmach D, Alders R. nd. Exploring hygiene practices, water supplies, sanitation and the occurrence of diarhoea in Central Tanzania: a mixed-method study. *J Prev Med Hygiene.* (under review).
79. World Bank. *Population, Total—Timor-Leste*. World Bank Group (2019). Available online at: https://data.worldbank.org/indicator/SP.POP.TOTL?locations=TL (accessed April 14, 2020).
80. Lopes M, Nesbitt H. Improving food security in Timor-Leste with higher yielding crop varieties. In *Paper presented at Australian Agricultural and Resource Economics Society Mini Symposium "Food Security in our Own Back-yard"*, Fremantle (2012).
81. da Costa MDJ, Lopes M, Ximenes A, Ferreira ADR, Spyckerelle L, Williams R, et al. Household food insecurity in Timor-Leste. *Food Security*. (2013) 5:83–94. doi: 10.1007/s12571-012-0228-6
82. General Directorate of Statistics (GDS), Ministry of Health (MoH), ICF. *Timor-Leste Demographic and Health Survey 2016*. Dili; Rockville, MD: GDS and ICF (2018).
83. Glazebrook D, Lopes M, da Costa MDJ, Ximenes A. Stocks and flows of household food supplies during the wet and dry seasons and food shortage period: a longitudinal case study among subsistence farmers in Aileu, Baucau. *Liquica and Manufahi districts, Timor-Leste*. Dili: Seeds of Life (2007).
84. Southeast Asian Ministers of Education Organization Regional Centre for Food and Nutrition (SEAMEO-RECFON). *Timor-Leste Food and Nutrition Survey 2013*. Final report. Jakarta: SEAMEO-RECFON (2015).
85. General Directorate of Statistics (GDS), United Nations Population Fund (UNFPA). *Population and Housing Census of Timor-Leste, 2010, Vol. 4, Suco Report*. Dili: GDS and UNFPA (2011).
86. Wong JT. *Exploringing the Links Between Village Poultry Health and Maternal and Child Diets and Nutrition in Timor-Leste* (Ph.D. thesis). University of Sydney, Sydney, NSW, Australia (2020).
87. Wikipedia. *Uganda*. Wikipedia (2020). Available online at: https://en.wikipedia.org/wiki/Uganda (accessed April 22, 2020).
88. UNICEF. *Situation Analysis of Children in Uganda.* UNICEF (2015). Available online at: https://www.unicef.org/uganda/what-we-do/nutrition (accessed April 22, 2020).
89. Catley A, Lotira R, Hopkins C. *Hidden Peaks: Women's Knowledge On The Seasonality And Root Causes Of Child Malnutrition In Karamoja, Uganda And Their Programming Preferences*. Kampala: Karamoja Resilience Support Unit; USAID/Uganda; UK aid and Irish Aid (2018). Available online at: https://karamojaresilience.org/publications/item/hidden-peaks-women-s-knowledge-on-the-seasonality-and-root-causes-of-child-malnutrition-in-karamoja-uganda-and-their-programming-preferences (accessed January 29, 2020).
90. Government of Uganda, UNICEF, UK aid, World Food Programme. *Food Security and Nutrition Assessment of the Karamoja Sub-region.* WFP and UNICEF (2018). Available online at: https://karamojaresilience.org/publications/item/food-security-and-nutrition-assessment-for-karamoja-sub-region-2 (accessed January 29, 2020).
91. Government of Vietnam. *Socialist Republic of Vietnam Government Portal: Overview on Vietnam Geography*. Government of Vietnam (2019). Available online at: http://www.gov.vn/portal/page/portal/English/TheSocialistRepublicOfVietnam/AboutVietnam (accessed April 27, 2020).
92. World Bank. *The World Bank In Vietnam: Overview*. World Bank Group (2019). Available online at: https://www.worldbank.org/en/country/vietnam/overview (accessed April 27, 2020).

93. FAO. *Family Farming Knowledge Platform: Vietnam*. FAO (2020). Available online at: http://www.fao.org/family-farming/countries/VNM/en/ (accessed April, 20 2020).
94. FAO. *Small Family Farms Country Factsheet: Vietnam*. Rome: Food and Agriculture Organisation of the United Nations (2018). Available online at: http://www.fao.org/3/i8358en/I8358EN.pdf (accessed December 11, 2020).
95. FAO. *Livestock Sector Brief: Vietnam. Livestock Information Sector Analysis and Policy Branch*. Rome: Food and Agriculture Organisation of the United Nations (2005).
96. Fermet-Quinet E, Hunt T, Stratton J. *OIE PVS Gap Analysis Report: Vietnam*. Paris: World Organisation for Animal Health (OIE) (2010).
97. Delabouglise A, Dao T, Truong D, Nguyen T, Nguyen NTX, Duboz R, et al. When private actors matter: information-sharing network and surveillance of highly pathogenic avian influenza in Vietnam. *Acta Tropica*. (2015) 147:38–44. doi: 10.1016/j.actatropica.2015.03.025
98. Delabouglise A, Antoine-Moussiaux N, Phan T, Dao D, Nguyen T, Truong B, et al. The perceived value of passive animal health surveillance: the case of highly pathogenic avian influenza in Vietnam. *Zoonoses Public Health*. (2016) 63:112–28. doi: 10.1111/zph.12212
99. FAO. Lessons from HPAI—A technical stocktaking of outputs, outcomes, best practices and lessons learned from the fight against highly pathogenic avian influenza in Asia 2005–2011. In *FAO Animal Production and Health*. Rome (2013).
100. O'Neill J. Tackling drug-resistant infections globally: final report and recommendations. In: *The Review on Antimicrobial Resistance*. London: Wellcome Trust and HM Government (2016).
101. Government of Vietnam. *Aide Memoire: Multi-stakeholder Engagement to Combat Antimicrobial Resistance in Viet Nam*. Hanoi: Government of Vietnam (2015).
102. Robinson TP, Bu DP, Carrique-Mas J, Fèvre EM, Gilbert M, Grace D, et al. Antibiotic resistance is the quintessential One Health issue. *Trans R Soc Trop Med Hyg*. (2016) 110:377–80. doi: 10.1093/trstmh/trw048
103. Cooper T, Grace D, Van De Fliert E, Nguyen-Viet H, Nguyen-Thanh B, Bui-Thi B, et al. Engaging with complexity for improved veterinary antimicrobial stewardship in Thai Nguyen, Vietnam. In: *Poster Prepared for the One Health and Antimicrobial Resistance Research Coordinating Workshop, 7–8 October 2019*. Hanoi (2019). Available online at: https://hdl.handle.net/10568/105863 (accessed December 11, 2020).
104. Truong BD, Doan PH, Tran D, Vinh K, Nguyen CV, Bach KT, et al. Assessment of drivers of antimicrobial usage in poultry farms in the Mekong Delta of Vietnam: a combined participatory epidemiology and Q-sorting approach. *Front Vet Sci*. (2019) 6:84. doi: 10.3389/fvets.2019.00084
105. Gardner E, Alders RG. Livestock risks and opportunities: Newcastle disease and avian influenza in Africa. *Planet@Risk*. (2014) 2:208–11.
106. Rushton J, Bruce M, Bellet C, Torgerson P, Shaw A, Marsh T, et al. Initiation of global burden of animal diseases programme. *Lancet*. (2018) 392:538–40. doi: 10.1016/S0140-6736(18)31472-7
107. FAO, OIE, WHO. *Memorandum of Understanding Between FAO, OIE and WHO Regarding Cooperation to Combat Health Risks at the Animal-Human-Ecosystems Interface in the Context of the "One Health" Approach and Including Antimicrobial Resistance*. Rome: Food and Agriculture Organisation of the United Nations (2018). Available online at: https://www.who.int/zoonoses/MoU-Tripartite-May-2018.pdf?ua=1 (accessed September 9, 2019).
108. FAO, IFAD, UNICEF, WFP, WHO. *The State of Food Security and Nutrition in the World 2019: Safeguarding Against Economic Slowdowns and Downturns*. Rome: Food and Agriculture Organisation of the United Nations (2019). Available online at: http://www.fao.org/3/ca5162en/ca5162en.pdf (accessed September 9, 2019).
109. Jost CC, Mariner JC, Roeder PL, Sawitri E, Macgregor-Skinner GJ Participatory epidemiology in disease surveillance and research. *Revue Scientifique et Technique-Office International des Epizooties*. (2007) 26:537–49. doi: 10.20506/rst.26.3.1765
110. CDC. *One Health*. US Centers for Disease Control and Prevention (2020). Available online at: https://www.cdc.gov/onehealth/index.html (accessed June 6, 2020).
111. Rockström J, Steffen W, Noone K, Persson Å, Chapin FS, Lambin EF, et al. A safe operating space for humanity. *Nature*. (2009) 461:472–5. doi: 10.1038/461472a
112. Raworth K. A doughnut for the anthropocene: humanity's compass in the 21st century. *Lancet Planet Health*. (2017) 1:e48–9. doi: 10.1016/S2542-5196(17)30028-1
113. Rampa F, Dekeyser K, Alders R, Dar O. *The Global Institutional Landscape of Food and Agriculture: How to Achieve SDG 2*. European Center for Development Policy Management and Chatham House (2019). Available online at: https://ecdpm.org/wp-content/uploads/Global-Institutional-Landscape-Food-Agriculture-How-To-Achieve-SDG2-ECDPM-Discussion-Paper-265-With-Chatham-House.pdf (accessed August 27, 2020).
114. Blackstock KL, Kelly GJ, Horsey BL. Developing and applying a framework to evaluate participatory research for sustainability. *Ecol Econ*. (2007) 60:726–42. doi: 10.1016/j.ecolecon.2006.05.014
115. Thomas P. *Introduction: challenges for participatory development in contemporary development practice. Development Bulletin 75*. The Development Studies Network, Australian National University (2013). p. 4–9. Available at: https://crawford.anu.edu.au/rmap/devnet/devnet/db-75.pdf (accessed July 9, 2020).
116. Brinkerhoff DW, Frazer S, McGregor-Mirghani L. *Policy Brief: Adapting to Learn and Learning to Adapt: Practical Insights from International Development Projects*. Research Triangle Park; RTI Press; RTI International (2018)
117. Hetherington E, Eggers M, Womoyi J, Hatfield J, Manyama M, Kutz S, et al. Participatory science and innovation for improved sanitation and hygiene: process and outcome evaluation of project SHINE, a school-based intervention in Rural Tanzania. *BMC Public Health*. (2017) 17:172. doi: 10.1186/s12889-017-4100-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

METHODS
published: 26 January 2021
doi: 10.3389/fvets.2020.546651

A Practical Introduction to Mechanistic Modeling of Disease Transmission in Veterinary Science

Carsten Kirkeby[1]*, *Victoria J. Brookes*[2,3], *Michael P. Ward*[4], *Salome Dürr*[5] and *Tariq Halasa*[1]

[1] Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Frederiksberg, Denmark, [2] School of Animal and Veterinary Sciences, Faculty of Science, Charles Sturt University, Wagga, NSW, Australia, [3] Graham Centre for Agricultural Innovation (Charles Sturt University and NSW Department of Primary Industries), Wagga, NSW, Australia, [4] Faculty of Veterinary Science, Sydney School of Veterinary Science, University of Sydney, Sydney, NSW, Australia, [5] Department of Clinical Research and Public Health, Veterinary Public Health Institute, University of Bern, Bern, Switzerland

Computer-based disease spread models are frequently used in veterinary science to simulate disease spread. They are used to predict the impacts of the disease, plan and assess surveillance, or control strategies, and provide insights about disease causation by comparing model outputs with real life data. There are many types of disease spread models, and here we present and describe the implementation of a particular type: individual-based models. Our aim is to provide a practical introduction to building individual-based disease spread models. We also introduce code examples with the goal to make these techniques more accessible to those who are new to the field. We describe the important steps in building such models before, during and after the programming stage, including model verification (to ensure that the model does what was intended), validation (to investigate whether the model results reflect the modeled system), and convergence analysis (to ensure models of endemic diseases are stable before outputs are collected). We also describe how sensitivity analysis can be used to assess the potential impact of uncertainty about model parameters. Finally, we provide an overview of some interesting recent developments in the field of disease spread models.

Keywords: simulation model, transmission model, disease dynamics, mechanistic model, disease model

OPEN ACCESS

Edited by:
Flavie Vial,
Animal and Plant Health Agency,
United Kingdom

Reviewed by:
Jörn Gethmann,
Friedrich-Loeffler-Institute, Germany
Amy Carr Kinsley,
University of Minnesota Twin Cities,
United States

***Correspondence:**
Carsten Kirkeby
Ckir@sund.ku.dk

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *29 March 2020*
Accepted: *21 December 2020*
Published: *26 January 2021*

Citation:
Kirkeby C, Brookes VJ, Ward MP, Dürr S and Halasa T (2021) A Practical Introduction to Mechanistic Modeling of Disease Transmission in Veterinary Science. Front. Vet. Sci. 7:546651. doi: 10.3389/fvets.2020.546651

INTRODUCTION

A disease spread model is a simplified representation of a real-life system of disease transmission. As defined by Lessler and Cummings (1), disease spread models (also known as mechanistic models of disease spread) include explicit hypotheses of the biological mechanisms that drive infection dynamics. Therefore, they differ from statistical models such as regression models. Disease spread models are motivated by a need to better understand the transmission dynamics of a disease, predict the spread of the disease in a population and its effects, and study how the spread can be influenced (including the evaluation of different strategies to improve surveillance and control of diseases). The quote, "all models are wrong, but some are useful," (2) is often stated because disease spread models are simpler than reality, but they generate information which is otherwise difficult to obtain (3). For example, experiments on disease transmission and control might insufficiently represent real-life disease ecology, or not be feasible due to high resource requirements (such as

time and monetary costs), or logistical and ethical constraints. In addition, observational studies of disease spread might not provide comparisons of the relevant control strategies, or not occur in the population of interest (e.g., transboundary diseases).

Models of disease transmission can represent diverse diseases, including bacterial and viral infections, as well as parasites and vector-borne diseases, in a range of host populations and environments, and at different scales (4). Disease spread models might identify critical elements and knowledge gaps by reconstructing a system using available knowledge (5). They can also be useful decision-making tools by simulating surveillance or control of a specific disease and comparing strategies in specific contexts, such as outbreak situations (6, 7). Models have also been used to inform outbreak preparedness [e.g., (8, 9)], and the control of endemic pathogens [e.g., (10–13)].

Here, we focus on modeling the spread of infectious diseases of animals in a range of contexts. The methods described are not unique to veterinary systems and are used in other disciplines such as ecology and human health. In particular, we focus on a class of model called individual-based models (IBMs). Mancy et al. (4) provide an in-depth discussion of the different motivations for developing disease spread models in ecology and animal health. They present a conceptual framework to guide model construction, focusing on the pre-modeling stage (model selection, establishing, and testing the theory). In building on Mancy et al. (4) our objectives are 3-fold; (1) to provide a practical introductory guide to the process of developing a mechanistic model of animal disease transmission using IBMs, aimed at researchers beginning in this field; (2) to describe important concepts before, during and after the programming stage of developing model of animal disease transmission; and (3) to provide practical examples of models, including code, in veterinary science. Thus, we provide a hands-on introduction to model building, and its use and challenges, for scientists starting to work on disease spread models.

METHODS

Definitions and Concepts

Before we describe the steps of model building in the context of IBMs, we briefly describe some key terms, concepts, and approaches applied in disease spread modeling. Terminology in this field can be inconsistent; for a list of terms and definitions used throughout this guide, see **Appendix 1**.

Terms Used in Disease Spread Modeling

Disease spread models simulate the transmission of an infectious disease between the disease hosts, who are modeled as *units of interest*. This unit is the smallest entity of the model and could be an individual animal (or part of it; for example, a quarter of the udder in a mastitis model), a group of animals, herds, or populations in regions or countries. The units of interest can be aggregated and modeled as proportions of the total population in each disease state (see below) at a given time, or modeled as individuals whose disease status is tracked through the disease states included in the model.

The simulated system includes time, making the model *dynamic*. Time can be modeled as a *continuous* or *discrete* process. In the latter a fixed time-interval is chosen and the model steps through each consecutive interval (time-step) and updates the numbers of units of interest in each disease state from the beginning to the end of the simulated period (for example, every day, for a year) or until the disease fades out. In contrast, if time is modeled as a continuous process, the rate of change in the relative numbers of units of interest in each disease state in the system is continuously modeled using differential equations.

For discrete time models, the length of a time step is designated by the modeler and depends on the disease dynamics, purpose of the model (for example, predictions in monthly time-steps might be useful for surveillance or disease control), the availability of data needed to parameterize the model (outbreak data might only be available on a yearly scale), and the time spent by an individual unit of interest in each disease state of the model (see below). Whilst daily time-steps are typical for most discrete disease-spread models (11), weekly (14) or biweekly [e.g., (15)], biannual (16), or even yearly time steps can be used [for example, when simulating long duration control programs, such as (13)].

A model can be *deterministic* or *stochastic*. A model is *stochastic* when there is variation in model outputs arising from the use of distributions to describe input parameters (rather than fixed values), or by allowing model events to occur as random processes (inherent stochasticity). See section "Modeling Disease Transmission" for illustration of the difference between deterministic and stochastic. The outputs from a stochastic model will vary every time the model is run. In contrast, outputs from deterministic models are consistent each time the model is run.

Disease spread models represent the dynamics of infection, or progression of the modeled units of interest through *disease states*, for instance *Susceptible (S)*, *Infectious* (*I*), and *Recovered (R)* states (an SIR model). In a susceptible state, a unit of interest has yet to be exposed to an infectious individual and infected (termed "effective contact"). Once effective contact has occurred, an individual is in an infectious state prior to transition to a recovered state (or death). This basic formulation can be expanded with other disease states; for example, an *Exposed (E)* state representing the latent period of the infection can occur prior to transitioning to the *Infectious (I)* state [for example, within-herd spread of FMD; (17)]. The modeled states are dependent on the natural history of the disease, the purpose and scale of the model, and the resolution of available data. For example, differentiation of clinical and subclinical infectious states can be included if the subclinical state is considered significant to spread given the scale of the model, or if clinical detection of the disease is an essential aspect in the model. In a model of rabies spread, the pre-infectious period of rabies was considered essential to include in a model in which the dog populations were small (18), and not considered necessary in a similar but larger-scale model of rabies spread in dog populations in Chad (19). We illustrate how the dynamics of infection as modeled in an SEIR model relate to the dynamics of disease (the observed states) in **Figure 1**.

The way in which the units of interest contact each other, or how they "mix," is a core component of a disease model. *Homogeneous* contact means that all the units have equal probability of contact with each other (no clustering). *Heterogeneous* contact means that the probability of contact between units of interest is not equal, hence clustering (spatial or related to other contact characteristics) exists in the population. Heterogeneous contact can be modeled by stratifying models into population groups (for example, by age or farm type), modeling contacts between units of interest according to a network structure, or modeling specific characteristics of units that influence contact [for example, furious rabies in dogs; (18)].

Modeling Approaches

Since Kermack and McKendrick first formulated the basic compartmental equation-based SIR model using differential equations in 1927 (20), numerous approaches to modeling disease transmission have been developed. For a comprehensive description of modeling approaches, see Mancy et al. (4). Briefly, models can be classified according to how the disease hosts are modeled (as individual units of interest, or as groups in which the proportion of units of interest in disease states are followed) and how contact occurs (the connectivity between units), then further differentiated on how time is modeled (discrete or continuous) and whether stochasticity is included.

Here, we focus on individual-based models [IBMs, or Individual-level models; Mahsin et al. (21)] in which individual units of interest are described and followed through the disease states. The units of interest in IBMs represent discrete entities (such as individuals or herds) and time steps are discrete.

An advantage of IBMs is that units of interest can be assigned their own properties that can influence disease transmission, detection or control. They are therefore useful to simulate heterogeneity in disease transmission between the units of interest. For example, in a model of foot-and-mouth disease (FMD), an individual herd might be predominantly either sheep or cattle, which might influence disease susceptibility and transmission at the herd level (22, 23). Agent-based models (ABMs) are a subset of IBMs in which contact—and hence disease transmission—is simulated between explicit pairs of individual units of interest. ABMs often include explicit movement of—and therefore, contacts between—individual units of interest, thus introducing contact heterogeneity in the population due to spatial variation (24). In an example in which rabies transmission was modeled, individual dogs were assigned specific roaming characteristics that influenced their contacts with other dogs (25). In a further example, heterogeneity of contacts between individuals was assigned using individuals' social network parameters (18, 19). Consequently, these models can have a high level of complexity, but also be computationally intensive (and consequently, relatively slow to implement and simulate).

If the unit of interest in an IBM is a group of individual animals (for example, herds), within-group disease spread can be modeled using an equation-based model with proportions of the unit of interest in disease-state compartments. In this case, specific individuals are not tracked. Such models are called nested models in ecological modeling (26).

Building an Individual-Based Model

Model building can be divided into three stages: pre-programming, programming and post-programming. These stages are common to all model types, and include different elements that should be considered (**Figure 2**). We describe the concepts associated with each stage in detail below (labeled according to **Figure 2**).

In **Appendix 2** (and https://github.com/ckirkeby/MDT), code examples are shown. We include code for a *difference equation model,* and a *differential equation model* (two model types not addressed in this article, but to enable the readers to compare the inputs and outputs with IBMs), and IBMs, for which we include examples of an *individual-based stochastic model (at herd level),* and an *individual-based stochastic model (at individual animal level).* We link the code for IBMs with each stage below.

Pre-programming Stage

Purpose

When designing a model, it is important to consider the research question to be investigated. This not only drives the type of model that might be appropriate, but also dictates the model outputs required by the end-user (27).

For example, whilst a model generally estimates the epidemiological consequences of the disease in terms of the number of infected individuals and epidemic duration, in the case of exotic diseases, the outputs could also be needed for contingency planning to improve surveillance and control; for example, identifying sentinel herds, culling capacity, or laboratory capacity [for example, (28, 29)]. In this case, it is essential to generate capacity-related data, such as the number of surveillance teams required, by including these parameters in the model. Similarly, if the purpose is to compare different surveillance strategies, sensitivity and specificity of tests used to detect disease need to be included (30, 31).

Evaluation and identification of optimal control strategies given a particular set of circumstances and constraints might also be a goal [for example (12, 17, 23, 32)]. This would require policy-specific knowledge to inform model processes, as well as data and knowledge of mechanisms to simulate control strategies. For example, to simulate vaccination, estimates of vaccination-specific parameters such as the number of individuals or herds vaccinated per day, vaccine efficacy, time required to order vaccine and perform vaccination could be included (9, 32). In addition to epidemiological metrics, the optimal control strategies could be defined according to economic outputs (33) such as in a bio-economic disease spread model [for example, (11)].

In the context of an IBM, the minimum inputs that must be included are a parameter to describe disease transmission (β; see later), and the number of individuals in each disease state. This will include at least one infectious individual as well as susceptible individuals (see code example, **Appendix 2**; https://github.com/ckirkeby/MDT). Additional parameters, such as the number of surveillance teams deployed, can be included as the model steps

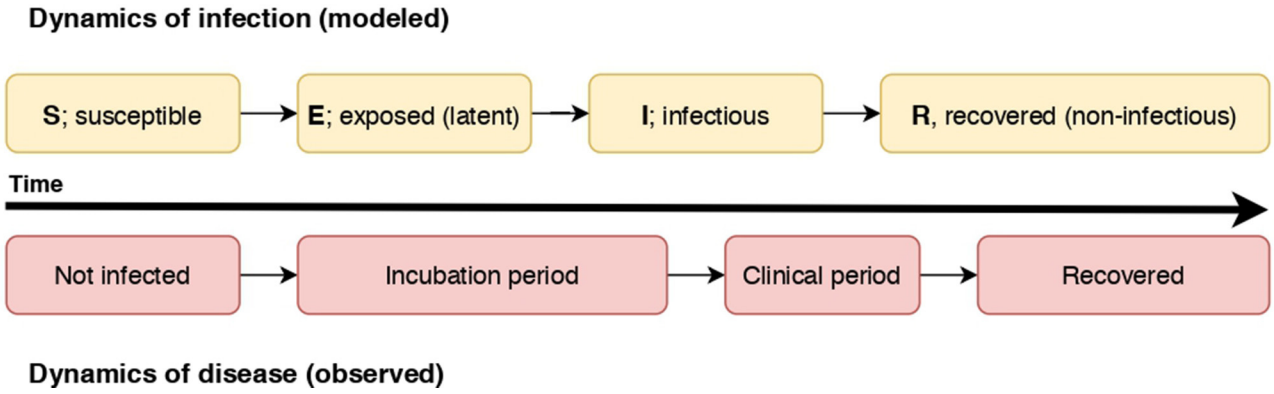


FIGURE 1 | A diagram illustrating the relationship of the dynamics of an SEIR infectious process and observed disease states in individuals. In this case, individuals become non-infectious prior to resolution of clinical signs.

through the discrete time intervals; for example, in response to trigger levels such as a threshold number of infected animals for disease detection.

Unit of Interest

The largest unit of interest is selected so the disease spread model sufficiently represents the true system. As described previously, this epidemiological unit of the model can range from individuals [e.g., (16)] or their parts [e.g., (12)] to sub- or entire populations (34).

The choice of epidemiologic unit of interest is highly dependent on the purpose of the model, the disease of concern and the data available to parameterize the model. In models in which disease spread needs to be captured at the individual animal level (for example, because disease detection or control is performed at this level), individual animals are modeled and followed. In the case of modeling the spread of an exotic disease in animals aggregated in herds, the herd might be a more realistic unit to model, because surveillance and decisions occur at the herd-level.

Practical programming considerations also influence the choice of this unit of interest. For example, it is more likely that individual animals as units of interest are computationally more challenging, and therefore, herds are often mire suitable to be the epidemiologic unit of interest (see also Section *Programming stage*). In some systems, there might be more than one unit of interest to be modeled, as in the case of vector-borne diseases—both the vector and the animal can be units of interest (35).

In **Appendix 2** we provide code examples of IBMs using different units of interest (also available online at https://github.com/ckirkeby/MDT).

System Knowledge, Complexity, and Data Availability

To create a model that is a sufficient representation of a real-life system, decisions need to be made about which known processes to include and exclude. This decision is bound to available information on the system. Such information is important to gather prior to model building to assess the level of uncertainty that is due to limited knowledge, how much data about the system is available, and the feasibility of delivering requested outputs. If essential data are missing to fulfill the designated purpose, options include collecting more data before modeling is initiated, re-specifying model complexity, or re-evaluating the model purpose. Following the principle of parsimony, a model should only be as complex as necessary to achieve the model purpose, thereby requiring the minimum number of assumptions (36).

Processes that should be considered include the population dynamics of the unit of interest (birth and death rate, and lifespan—this is usually based on age, or in the case of a livestock production system, this could be parity), migration of individual units in and out of the system, the contact patterns between the units and the production system of the modeled population (for example, milk or beef production), if this is relevant. It also includes knowledge of the epidemiology of the disease to be modeled, such as the relevant disease states and their durations, the modes of transmission of the causative pathogen (for example, whether or not airborne spread is an essential pathway of transmission) and how the disease develops in the individuals.

Model Type Selection

Model specification (units of interest, disease, and system dynamics and how they are modeled—for example, discrete vs. continuous time and deterministic vs. stochastic) is typically an iterative process and is re-examined as data gathering for parameterization occurs (**Figure 2**, section Documentation and Communication). If data about population dynamics, disease dynamics and the system in which disease occurs are available at an individual level, and modeling at this level of detail and heterogeneity is considered valuable (for example, if the population is small or heterogeneity of the system is considered an important feature of disease transmission), an IBM is likely suitable. Otherwise, other model types can be considered (4).

In **Figure 3** we show the difference in output between a deterministic and a stochastic model.

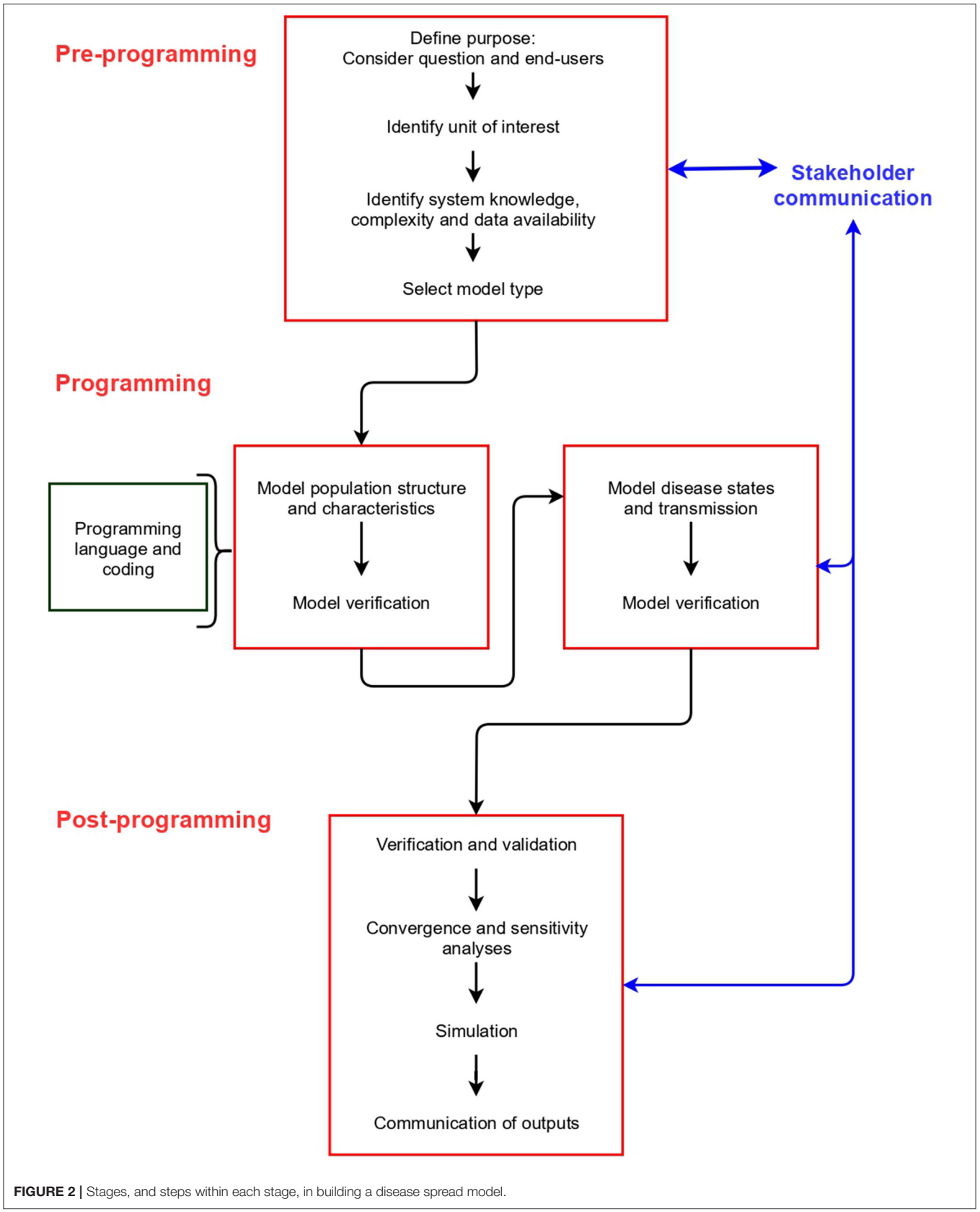


FIGURE 2 | Stages, and steps within each stage, in building a disease spread model.

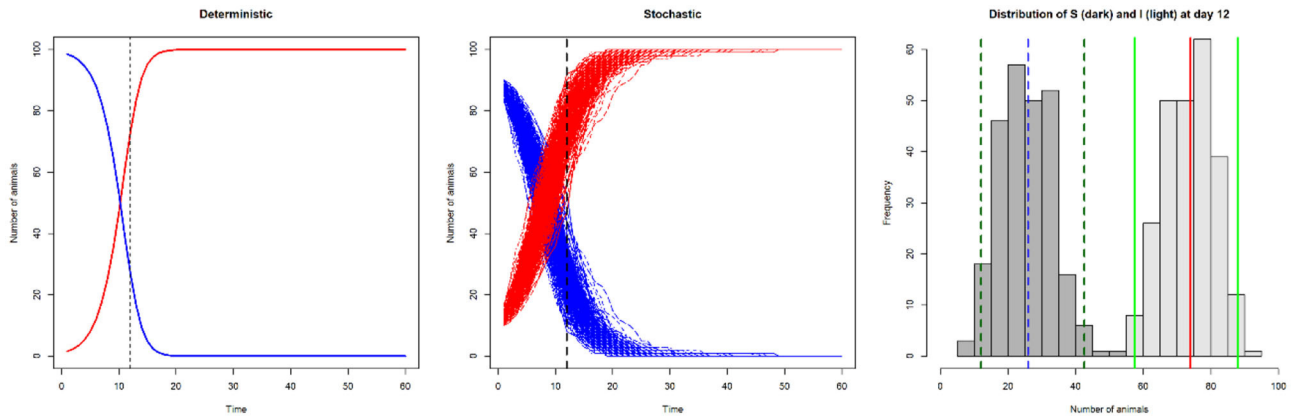


FIGURE 3 | Line plots of number of susceptible (S; blue) and infected (I; red) animals in a model with SI disease dynamics, illustrating deterministic output (left: number of infected and susceptible animals on day 12 is 73 and 27, respectively), and stochastic output (centre). The histogram illustrates the variation in the number of susceptible (dark gray; median 26 [blue dashed line], 95% CI 11–42 [dark green dashed lines]) and infected animals (light gray; median 74 [red line], 95% CI 58–89 [light green lines]) from the stochastic output at day 12.

Programming Stage

Programming Language and Coding

Programming languages can be classified in many ways—such as whether interpreted directly or compiled (running one single line of code at a time, rather than all the code has to be run together; for example, Python and R vs. C++ and Fortran, respectively); and whether they are "high" or "low" level languages. This latter classification refers to the machine-readability of the language; many languages used in the context of disease modeling can be considered high-level (for example, Java, C++, R, and Python).

In general, programs written using high-level languages require more memory space but are more readable by a human, and therefore more accessible to people without detailed programming knowledge. Programs written using low-level languages (e.g., Assembly language) can better utilize hardware specific features. These programs require a high level of knowledge to write and maintain. They can be hardware-dependent making them less portable between computer architectures.

Features resulting from language classification are not always exclusive; with many factors affecting the overall performance and efficiency of a program. For example, a complex "real-world" program written in a more user-friendly and high-level language with a modern optimizing compiler can produce highly efficient machine code with excellent performance. The result is likely to outperform an equivalent program hand-written in the less user-friendly, low-level Assembly language converted to machine code via an assembler. Advances in computational power and improvements in system architecture enable the horizontal scaling of models by running processes in parallel across multiple cores to reduce "wall time" (the time taken to complete a simulation).

Focusing on final run speed also ignores the concept of overall programming productivity. Programming in some languages is more challenging and less accessible to the research team, which increases the time required for programming. An increasing number of researchers use the free software R (37), which is a statistical programming language suitable for building many model types, including equation-based [for example, (38)] and individual-based models [for example, (11, 32)]. There are many packages available for languages such as R, and they are well-supported and maintained by R's open-source community, which allows the team to focus on modeling the system and the disease.

In regards to code programming, we highly recommend that modelers annotate their code during modeling with detailed descriptions of each part of the code. For a description of good practice in animal health modeling, see EFSA (39). Annotation assists the modeler to remember the function of each line of code, and also facilitates use of the model by others. Following publication of a study, it is a requirement of many journals that the code be made available to readers. Version control such as git (https://git-scm.com, accessed 10/09/2019) is a very valuable tool so that modelers can easily track changes in the code, and view previous versions (branches) of the model. This is of particular value when more than one modeler is involved in the project or when published code is used by other researchers. Locally, version control can be as simple as saving the script in a new file named with the specific day it is changed. We also highly recommend that during the programming process, each line or chunk of co-de should be executed with fictitious inputs to check for errors (debugging). This is part of the model verification (see section *Model Verification and Validation* for more details).

Modeling the Population Structure and Characteristics

Initially when constructing an IBM, the host population dynamics are modeled as the "background" for the disease dynamics. For example, a model of canine rabies spread requires a population of dogs or a foot-and-mouth disease model the population of cloven-hoofed animals. An understanding of the population of interest's demographics are critical. Whilst demographic data for livestock populations can often be gained from government or industry sources, it might be necessary

to conduct studies of other populations (such as companion animals) prior to modeling to for example determine age structure and birth and death rates (40).

The population dynamics are linked to the disease model; for example, newborns can be susceptible, infected or immune (see section *Modeling disease transmission*). Also, characteristics can be allocated to the units of interest in case they influence disease transmission. In an example of Johne's disease (paratuberculosis) transmission, individual cattle or herds are modeled, and characteristics, such as individuals' milk production and lactation duration, are included because these characteristics influence disease spread [e.g., (11, 41)].

In disease spread models, it can be important to include a spatial component to the population to allow spatio-temporal modeling of disease transmission (see section *Modeling disease transmission*). This can be realized by using geolocations of the units of interest, e.g., farms, as a feature of the population structure [e.g., (17, 42)]. Spatio-temporal modeling could also represent population structures other than farms, as in the case of modeling spatio-temporal distributions of vectors that transmit bluetongue virus (43), or in the location of dog's residence in a rabies transmission model (44).

Once the background structure of the disease dynamic system has been modeled, it should be verified and tested (see sections *Model verification and validation*) before disease transmission is added to the model. This is to ensure that the model simulates the system with sufficient accuracy, as well as to determine computing requirements such as the number of iterations required for burn-in (see section *Modeling disease transmission*).

Modeling Disease States

As discussed previously, each stage of disease in the transmission model should reflect a -state during the course of infection in the modeled system. In the simplest framework, an *SI* model with two, mutually exclusive disease states; *Susceptible (S)* and *Infectious (I)*, all individuals in the model are assigned to either *S* or *I* (see code examples in **Appendix 2**; https://github.com/ckirkeby/MDT). For each simulated time step, each individual has a probability of acquiring infection and thus transitioning from *S* to *I*, depending on the contact pattern between individuals and the disease transmission rate given a contact. In the case of the *SI* model, there is no probability of individuals returning to the *S* state. In the case that animals can recover from the disease, the model becomes an *SIS* model in which infectious individuals return to the *S* state. The transmission from *I* to *S* is quantified by the recovery rate (see below, in the context of an SIR model), which can be influenced by self-recovery or by treatment. The recovery rate is thus a probability of recovering during each time step. Recovery rates must be estimated from epidemiological studies on the duration of infection. This duration of infection can either be modeled as a fixed timespan, i.e., a fixed number of days can be assigned to it, or as a distribution, after which it will revert to the *S* state.

Another common framework is the SIR model (see the code example in **Appendix 2**, https://github.com/ckirkeby/MDT), in which the infectious individuals can enter the *Recovered* (*R*) state – which represents either "recovery" (and resistance to infection) or "removal" from the population; for example, in the case of a rabies model, infected dogs always die and therefore are removed. The transition from *I* to *R* is also modeled via a recovery rate (denoted as "r" in the code example). Following this logic, the disease transmission framework can be further extended dependent on the disease; for example, by introducing an *Exposed (E)* state for latently infected individuals before progressing to the *I* state. As previously mentioned, even if some disease states occur in reality, it is not always useful or necessary to represent them in the model.

In the case of modeling endemic diseases, once the population and disease dynamics frameworks are modeled, an IBM might need to be simulated for enough time steps to reach a stable prevalence ("burn-in" period; the number of time steps for the population characteristics and the disease prevalence to stabilize). When such a model is used to assess control strategies, these strategies are usually implemented after the burn-in period, when a stable situation has been reached.

Modeling Disease Transmission

The process of disease transmission is the core dynamic process in the model. Generally, transmission can be considered as either direct (from host to host) or indirect, for example via the environment or vector transmitted (45). It can also be dependent on model features that increase contact heterogeneity; for example, some models are spatially explicit and the probability of transmission varies according to distance, mimicking a system in which transmission varies with spatial location (46).

Since disease transmission is the core process in a disease transmission model, we guide the reader through the foundation of this in the context of an IBM, such as those shown in code in **Appendix 2** (https://github.com/ckirkeby/MDT). In the case of direct transmission, we first describe β, a parameter that underpins the modeling of disease transmission in equation based models, and then we describe how this parameter can be used in IBMs (47). Beta is defined as the per capita rate at which two specific individuals come into effective contact per unit time [sometimes called the transmission rate; Vynnycky and White (48)]. An effective contact is one which is sufficient for disease transmission to occur. This effective contact rate, β, comprises a contact rate between individuals (C), and the probability of transmission per contact (P):

$$\beta = C \cdot P \tag{1}$$

The contact rate C in the above equation is defined per unit time, and is fundamentally different between density-dependent or frequency-dependent transmitted diseases (49–51). In density-dependent transmission, the greater the density of individuals, the greater the probability of contact per unit time (52):

$$\frac{dI}{dt} = \beta \cdot S \cdot I \tag{2}$$

where dI/dt is the rate of new infections per unit time t, β is the effective contact rate, and S and I are the number of susceptible and infected individuals, respectively.

In frequency-dependent transmission, the rate of new infections per unit time, dI/dt, is independent of the density of individuals in the population (N):

$$\frac{dI}{dt} = \beta' \cdot \frac{S \cdot I}{N} \quad (3)$$

where S and I are the same as in Equation 2, but β' is not equivalent to β in Equation 2 due to the underlying difference between the contact rates (C) of these two types of transmission. The difference between these two types of transmission is demonstrated in a study of mange in a fox population in the UK, in which researchers compared density and frequency dependent transmission and found that mange transmission was most likely frequency dependent in this population (53).

As an example of a method to allow a random process of becoming infected that can be used at each time step in an IBM, we extend Equation 3 to calculate a probability of infection per susceptible individual, *P(S)*, so each individual can be separately subjected to a Bernoulli process of becoming infected (54):

$$P(S) = 1 - \exp(-\beta' \cdot \frac{S \cdot I}{N}) \quad (4)$$

with the same notation as for Equation 2, and N is the total number of individuals in the modeled population. If β is fixed, then the probability of infection for all susceptible individuals is the same (for all individuals and all simulated time), and assumes homogeneity of transmission in the population. In IBMs, β may vary from one individual to another representing the susceptibility and infectiousness of the individual, thus representing natural heterogeneity in transmission. This could be driven by a lower probability of infection as a result of, for instance, vaccination or due to different contact rates between individuals.

The R code examples demonstrate this type of transmission in **Appendix 2** (https://github.com/ckirkeby/MDT). In this way, the infection pressure is scaled to the proportion of the population that are infected within each time step, i.e., I changes over time, whereas β and N (within a closed system) remain constant. The infection process is dynamic because the *P(S)* changes over time with changing numbers of I in the population (assuming a fixed N and β).

As mentioned at the start of this section, it is possible to consider the spatial structure of the underlying demography and define the probability of effective contact per time step for a susceptible unit of interest dependent on its distance from infectious units in the model. For this approach, distance kernels can be built from which the probability of effective contact can be drawn (such as used in 8, 23). This spatially dependent contact rate can be combined with information on the frequency of contacts between units of interest. For example, the frequency of potential contacts between herds may not only depend on the distance between them, but also on the frequency of movements between herds, which in turn may depend on the herd types (55, 56).

When appropriate knowledge and data are available, the contact structure of a population can be based on a social network (18, 57). A heterogeneous herd contact structure between groups of animals (for example, calves and heifers) and homogenous contacts within animal groups might also be described (11, 12).

There are also several ways to simulate indirect (environmental) disease transmission. It can be similarly spatially dependent as described for the direct transmission, or simulated as a fixed transmission probability:

$$P(S) = 1 - \exp(-\beta_i) \quad (5)$$

Here, *P(S)* is the probability of infection of a susceptible individual S, and β_i is the indirect disease transmission rate. This fixed transmission rate can be based on a stable baseline infection pressure, or more variable, such as bacteria from infected individuals shed over time in the environment (11).

When disease transmission occurs through both direct and indirect contacts, a combination of both of these direct and indirect pathways can be used (12).

In **Figure 4** we show an example of an SI model in which the transmission rate, β, is varied.

Post-programming Stage

Model Verification and Validation

Model verification and validation is essential to ensure that model concepts, programming and outputs are reliable, accurate, and representative for the modeled system (27, 58). Model verification ensures that model code and the conceptual framework are implemented correctly. Verification is also called computerized model verification, internal validation, or conceptual validation (58). Several methods can be used for model verification, including: (1) The rationalism method, in which several scenarios are simulated with different inputs, and outputs are compared to determine whether the changes in outputs are rational given the changes in the inputs (sensitivity analysis, see below); (2) The tracing method, in which individuals or other units of interest are followed through the different time steps and checked that they behave as expected; and (3) The face validitation method, in which an expert is asked to evaluate the outputs or even the code to verify the credibility of the model.

Model validation (also called external or operational validation) ensures that the model predictions have a satisfactory range of accuracy in relation to the actual behavior of the modeled system in real life (adapted from 54). Real-life data (i.e., empirical outbreak data) is needed to fully execute this process. To our knowledge, few models in veterinary science have been externally validated (59–61). This is usually due to the high associated costs or ethics of obtaining such data, and the complexity of the modeled systems. If empirical outbreak data are lacking from the setting in which the model was built and applied—such as in the case of exotic diseases and regions with historical disease freedom—then validation options might include either adapting the model to a region where data are available, or using previous outbreak data. For example, historical data from the last Swiss FMD outbreak was used to validate a current FMD model for Switzerland (61).

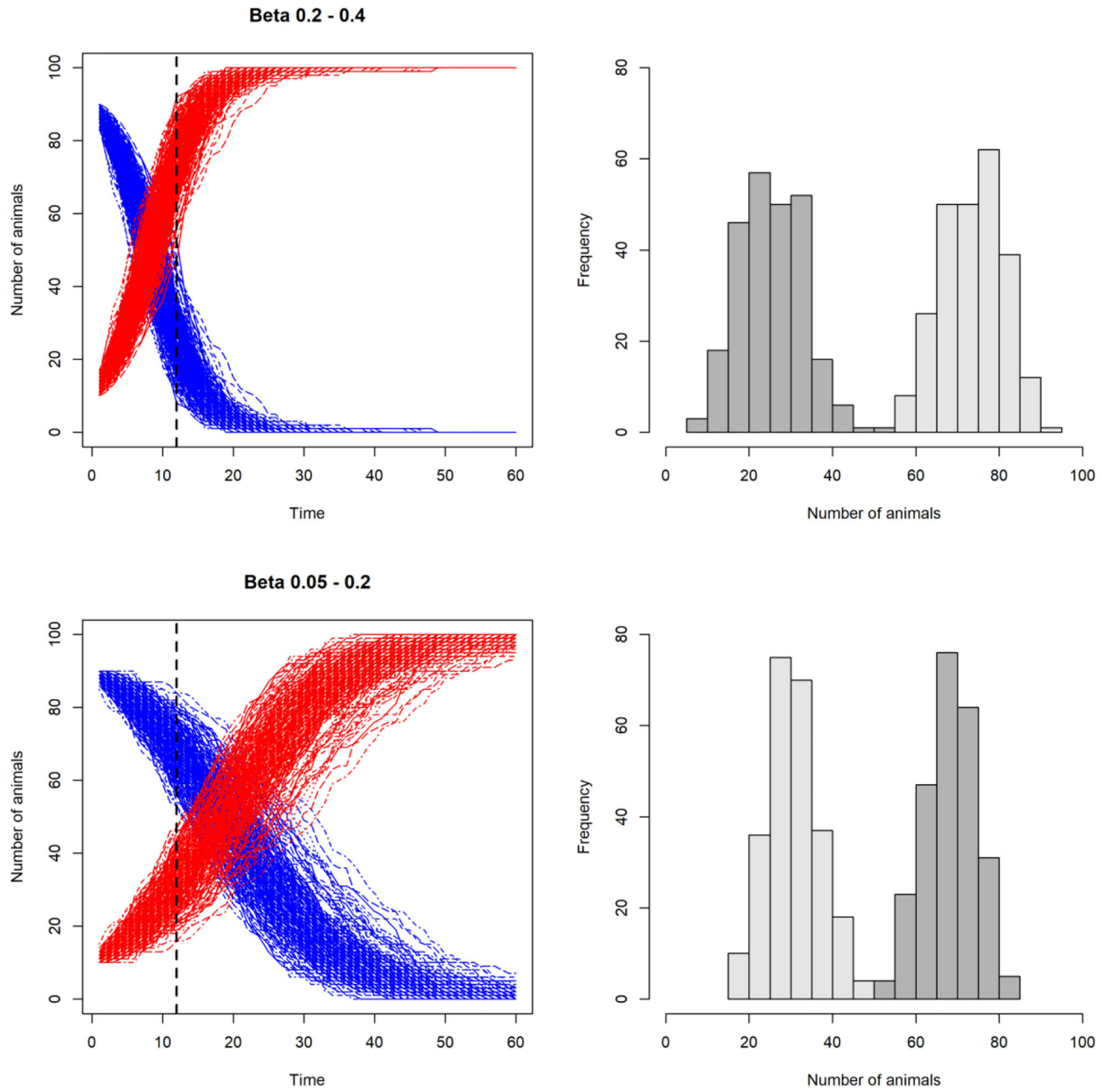


FIGURE 4 | Line plots illustrating the effect of varying the transmission rate, β, on the number of susceptible (*S*) and infected (*I*) animals in a stochastic model with SI disease dynamics. In the two upper plots, β is higher than the two lower plots. This results in higher number of infected than susceptible at day 12 in the upper scenario compared to the lower scenario. In the histograms on the right, the resulting distributions of susceptible (dark gray) and infected (light gray) individuals are shown. Note that β is not kept constant, but varied for each iteration, incorporating uncertainty around this key parameter.

Convergence Analysis

Convergence analysis assesses the repeatability of the outputs based on the number of iterations (repetitions) the model is simulated, and is conducted before final model simulations. Above a given threshold of simulations, the output statistics should be independent of the number of model iterations. This stability can be checked by ensuring that the variance of the outputs of interest (for example, the number of infected individuals or epidemic duration) is stable. A commonly used approach is to visualize the change in the variance when increasing the number of iterations (62), or to use thresholds of the coefficient of variance as a decision metric (9, 18, 63).

We have included an example of how to determine convergence of a model in **Appendix 2** (https://github.com/ckirkeby/MDT).

Sensitivity Analysis

Sensitivity analysis is essential to understand and examine the robustness of model predictions to changes in input parameter values, model structure and processes (64). Sensitivity analysis can be used to identify parameters and processes that have a major influence on model predictions; therefore, the values of these parameters—and the way in which processes are modeled—must be certain enough to produce model predictions acceptable to the end-user.

During sensitivity analysis, the behavior of the model and the outputs of interest are examined when the model or its parameters are varied. There are different ways to approach sensitivity analysis. Sensitivity analysis of input parameters can be assessed by changing input values within a specified range (local sensitivity analysis) or the entire parameter space (global sensitivity analysis) to examine the impact of these changes on model outputs. The influence of parameters can also be examined singly (one-at-a-time sensitivity analysis) or in combination with other parameters [for example, a "Sobol" sensitivity analysis, (65)]. Sensitivity analysis can also be implemented by modeling a specific process in alternative ways to examine the impact of this process on model predictions (this is sometimes referred to as structural sensitivity analysis).

The simplest method of sensitivity analysis of input parameters is one-at-a-time perturbations (66). However, this does not allow assessment of the sensitivity of the model output to changes in combinations of other parameter values' change. Many more methods exist and have been used in the context of IBMs (10, 66, 67); a complete review is beyond the scope of this article.

We have included code in **Appendix 2** to conduct a simple sensitivity analysis on a model parameter (also available online at https://github.com/ckirkeby/MDT).

Presentation of Model Outputs

Presentation of clear results that deliver project requirements is an important element for transparent communication of the model outputs. This should already be reflected and incorporated during the design stage. Deterministic models provide single value outputs (without variation), whereas stochastic models provide distributions of outputs. Thus, when results from stochastic models are presented, it is essential to not only show median or mean values, but also the variation around these values; for example, using boxplots or histograms. From a disease spread model, outputs usually include the number of infected units of interest and the epidemic duration. Other outputs can also include the number of units of interest under control (culled, vaccinated, or banned in movements), economic metrics in case of a bio-economic model, predicted changes in production (such as milk yield or growth rates), or maps from spatially-explicit models.

Documentation and Communication

Good documentation is essential to enable reproducibility of the model, communication of model outcomes, and comparison between different models. Standardized protocols for disease spread model documentation have been developed, such as the ODD (Overview, Design concepts, and Details) (68) and TRACE (69) and can be used to communicate models in scientific publications.

At all stages of model design, development and implementation, communication should be maintained with relevant stakeholders. These will include the end-users of the model, but can also include experts for the specific disease and system modeled, and those that are funding model development and implementation. Comprehensive communication at all stages ensures that the model focus remains on the defined purpose so that useful information is provided to the end-users, or that the end-user can adapt the model according to specific needs during the modeling process.

Recent Developments

Recent developments in disease spread models used in veterinary science include the development of models that model more than one disease. Mostert et al. (70) present a bio-economic stochastic dynamic model that simulates subclinical and clinical ketosis, mastitis, metritis, displaced abomasum, and lameness in dairy cattle. In intense production systems, such as in the dairy sector, it is an advantage to evaluate the impact of several diseases concurrently, to optimize management strategies. Inclusion of economic impacts and the economics of disease mitigation in these models facilitates broader use, in addition to improving animal welfare.

Many populations can also be captured in one model. One example is the trend for models of vector-borne diseases (which we have not covered here, and introduces at least one more population, the vector, into the model).

Ensemble modeling is a relatively new approach in veterinary epidemiology (71). Decisions on how to respond to an incursion of FMD virus in a previously disease-free country are complex and several models of FMD spread have been developed and applied. These vary in their disease processes modeled, assumptions made and parameterization. For any set of inputs, outputs from these various models are plausible. Variability in model outputs can be valuable because these are likely to include the range of realizations that could be observed during an FMD outbreak. A method of reconciling variability—borrowed from fields such as meteorology, climate-change science and medical science—has recently been applied to this situation. Using outputs from six different models which simulated the spread of FMD in the Midlands and Wales areas of the United Kingdom in 2001, Webb et al. (71) applied a Bayesian Reliability Ensemble Average (BREA) method to integrate outputs regarding outbreak duration and two control methods. The BREA method determines the weights applied to each model output based on agreement with observed data (bias criterion) and consensus between models (convergence criterion). The latter was used by Webb et al. (71) and their case study highlights the potential of ensemble modeling to reduce the uncertainty of outputs from individual models, thus improving decision-making.

CONCLUSIONS AND RECOMMENDATIONS

We emphasize two well-known, key axioms: (1). disease spread models are simplified representations of real-life systems so that "all models are wrong, but some are useful" (2), and (2). model outputs can only be as accurate as model inputs allow.

Model simplification is often driven by data availability; therefore, full use of any available data is recommended. However, when considering whether more data should be collected or how a process should be modeled, we note that highly detailed models (more complex processes with more parameters, such as IBMs) can produce output that might be less generalizable than more simplified models. In addition, the output from more simplified models might adequately predict the essential components of disease transmission needed to achieve the end-users' objectives. This presents modelers with dilemmas: a highly detailed model is not necessarily less "wrong" or more "useful" than a simplified model. Whilst the steps of model verification, validation, and sensitivity analysis can help avoid too much or too little simplification, we recommend that particularly during the design phase, modelers focus on development of the simplest model to achieve useful output—whilst we focus on an introduction to modeling using IBMs, we do not suggest that they are the foundation of modeling approaches.

Communication between end-users and modelers about the value and assumptions of a model is critical. We therefore recommend that modelers and end-users, wherever possible, establish a framework for communication about modeling objectives, the need for verification, validation, and sensitivity analysis, and application of model outputs to ensure optimal use of simulation modeling, to improve animal health, welfare, and production.

DATA AVAILABILITY STATEMENT

The original contributions generated for the study are included in the article/**Supplementary Materials**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

CK wrote the first draft of the manuscript. All authors participated in writing the manuscript.

ACKNOWLEDGMENTS

This article was initiated during a special course in simulation modeling held at the University of Sydney in June 2019, funded by the University of Sydney and University of Copenhagen Partnership Collaboration Awards 2019. We thank David Brookes for his advice on information technology.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.546651/full#supplementary-material

REFERENCES

1. Lessler J, Cummings DAT. Mechanistic models of infectious disease and their impact on public health. *Am J Epidem*. (2016) 183:415–22. doi: 10.1093/aje/kww021
2. Box GE. Science and statistics. *J American Stat Assoc*. (1976) 71:791–9. doi: 10.1080/01621459.1976.10480949
3. Holmdahl I, Buckee C. Wrong but useful—what Covid-19 epidemiologic models can and cannot tell us. *N Engl J Med*. (2020) 383:303–5. doi: 10.1056/NEJMp2016822
4. Mancy R, Brock PM, Kao RR. An integrated framework for process-driven model construction in disease ecology and animal health. *F Vet Sci*. (2017) 4:155. doi: 10.3389/fvets.2017.00155
5. Singer A, Salman M, Thulke HH. Reviewing model application to support animal health decision making. *Prev Vet Med*. (2011) 99:60–7. doi: 10.1016/j.prevetmed.2011.01.004
6. Zinsstag J, Dürr S, Penny MA, Mindekem R, Roth F, Gonzalez SM, et al. Transmission dynamics and economics of rabies control in dogs and humans in an African city. *PNAS*. (2009) 106:14996–5001. doi: 10.1073/pnas.0904740106
7. Thulke H-H, Lange M, Tratalos JA, Clegg TA, McGrath G, O'Grady L, et al. Eradicating BVD, reviewing Irish programme data and model predictions to support prospective decision making. *Prev Vet Med*. (2018) 150:151–61. doi: 10.1016/j.prevetmed.2017.11.017
8. Halasa T, Boklund A, Bøtner A, Mortensen S, Kjær LJ. Simulation of transmission and persistence of African swine fever in wild boar in Denmark. *Prev Vet Med*. (2019) 167:68–79. doi: 10.1016/j.prevetmed.2019.03.028
9. Dürr S, Ward MP. Development of a novel rabies simulation model for application in a non-endemic environment. *PLoS Negl Trop Dis*. (2015) 9:e0003876. doi: 10.1371/journal.pntd.0003876
10. Brookes VJ, Jordan D, Davis S, Ward MP, Heller J. Saltelli global sensitivity analysis and simulation modelling to identify intervention strategies to reduce the prevalence of *Escherichia coli* o157 contaminated beef carcasses. *PLoS ONE*. (2015) 10:e0146016. doi: 10.1371/journal.pone.0146016
11. Kirkeby C, Græsbøll K, Nielsen SS, Christiansen LE, Toft N, Rattenborg E, et al. Simulating the epidemiological and economic impact of paratuberculosis control actions in dairy cattle. *Fron Vet Sci*. (2016) 3:90. doi: 10.3389/fvets.2016.00090
12. Gussmann M, Kirkeby C, Græsbøll K, Farre M, Halasa T. A strain-, cow-, and herd-specific bio-economic simulation model of intramammary infections in dairy cattle herds. *J Theor Biol*. (2018) 449:83–93. doi: 10.1016/j.jtbi.2018.04.022
13. Zingg D, Steinbach S, Kuhlgatz C, Rediger M, Schüpbach-Regula G, Aepli M, et al. Epidemiological and economic evaluation of alternative on-farm management scenarios for ovine footrot in Switzerland. *Front Vet Sci*. (2017) 4:1–13. doi: 10.3389/fvets.2017.00070
14. Kudahl AB, Østergaard S, Sørensen JT, Nielsen SS. A stochastic model simulating paratuberculosis in a dairy herd. *Prev Vet Med*. (2007) 78:97–117. doi: 10.1016/j.prevetmed.2006.05.015
15. Halasa T, Nielen M, van Werven T, Hogeveen H. A simulation model to calculate costs and benefits of dry period interventions in dairy cattle. *Livestock Sci*. (2010) 129:80–7. doi: 10.1016/j.livsci.2010.01.009
16. Groenendaal H, Nielen M, Jalvingh AW, Horst SH, Galligan DT, Hesselink JW. A simulation of Johne's disease control. *Prev Vet Med*. (2002) 54:225–45. doi: 10.1016/S0167-5877(02)00027-2
17. Backer JA, Hagenaars TJ, Nodelijk G, van Roermund HJW. Vaccination against foot-and-mouth disease I: epidemiological consequences. *Prev Vet Med*. (2012) 107:27–40. doi: 10.1016/j.prevetmed.2012.05.012

18. Brookes VJ, Dürr S, Ward MP. Rabies-induced behavioural changes are key to rabies persistence in dog populations: investigation using a network-based model. *PLoS Negl Trop Dis*. (2019) 13:e0007739. doi: 10.1371/journal.pntd.0007739
19. Laager M, Mbilo C, Madaye,E. A., Naminou A, Léchenne M, Tschopp A, et al. The importance of dog population contact network structures in rabies transmission. *PLOS Negl Trop Dis*. (2018) 12:e0006680. doi: 10.1371/journal.pntd.0006680
20. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proce R Soc London Series A*. (1927) 115:700–21. doi: 10.1098/rspa.1927.0118
21. Mahsin M, Deardon R, Brown P. Geographically-dependent individual-level models for infectious diseases transmission. *Biostatistics*, 21. doi: 10.1093/biostatistics/kxaa009
22. Mardones F, Perez A, Sanchez J, Alkhamis M, Carpenter T. Parameterization of the duration of infection stages of serotype O foot-and-mouth disease virus: an analytical review and meta-analysis with application to simulation models. *Vet Res*. (2010) 41:45. doi: 10.1051/vetres/2010017
23. Ward MP, Highfield LD, Vongseng P, Garner MG. Simulation of foot-and-mouth disease spread within an integrated livestock system in Texas, USA. *Prev Vet Med*. (2009) 88:286–97. doi: 10.1016/j.prevetmed.2008.12.006
24. Tang W, Bennett DA. Agent-based modeling of animal movement: a review. *Geography Compass*. (2010) 4:682–700. doi: 10.1111/j.1749-8198.2010.00337.x
25. Hudson EG, Brookes VJ, Ward MP, Dürr S. Using roaming behaviours of dogs to estimate contact rates: the predicted effect on rabies spread. *Epidemiolo Infect*. (2019) 147:e135. doi: 10.1017/S0950268819000189
26. Mideo N, Alizon S, Day T. Linking within-and between-host dynamics in the evolutionary epidemiology of infectious diseases. *Trends Ecol Evol*. (2008) 23:511–7. doi: 10.1016/j.tree.2008.05.009
27. Garner MG, Hamilton SA. Principles of epidemiological modelling. *Rev Sci Tech*. (2011) 30:407–16. doi: 10.20506/rst.30.2.2045
28. Garner MG, Bombarderi N, Cozens M, Conway ML, Wright T, Paskin R, et al. Estimating resource requirements to staff a response to a medium to large outbreak of foot and mouth disease in Australia. *Transbound Emerg Dis*. (2016) 63:e109–21. doi: 10.1111/tbed.12239
29. Boklund A, Halasa THB, Christiansen LE, Willeberg P, Enøe C. Simulated effects of introducing emergency vaccination or depopulation during FMD outbreaks in Denmark. In: *EuFMD Meeting 2012: Open Session*, Jerez (2012).
30. Nusser SM, Clark WR, Otis DL, Huang L. Sampling considerations for disease surveillance in wildlife populations. *J Wildl Manage*. (2008) 72:52–60. doi: 10.2193/2007-317
31. Halasa T, Boklund A. The impact of resources for clinical surveillance on the control of a hypothetical foot-and-mouth disease epidemic in Denmark. *PLoS ONE*. (2014) 9:e102480. doi: 10.1371/journal.pone.0102480
32. Dürr S, Fasel-Clemenz C, Thür B, Schwermer H, Perler L, Doherr MG, et al. Evaluation of the benefit of emergency vaccination in a foot-and-mouth disease free country with low livestock density. *Prev Vet Med*. (2014) 113:34–46. doi: 10.1016/j.prevetmed.2013.10.015
33. Gethmann J, Probst C, Bassett J, Blunk P, Hövel P, Conraths FJ. An epidemiological and economic simulation model to evaluate strategies for the control of bovine virus diarrhea in Germany. *Front Vet Sci*. (2019) 6:406. doi: 10.3389/fvets.2019.00406
34. Kopec JA, Finès P, Manuel DG, Buckeridge DL, Flanagan WM, Oderkirk J, et al. A stochastic model simulating paratuberculosis in a dairy herd. *Prev Vet Med*. (2007) 78" 97–117.
35. Græsbøll K, Bødker R, Enøe C, Christiansen LE. Simulating spread of bluetongue virus by flying vectors between hosts on pasture. *Sci Rep*. (2012) 2:863. doi: 10.1038/srep00863
36. Lilien GL. Model relativism: a situational approach to model building. *Interfaces*. (1975) 5:11–18. doi: 10.1287/inte.5.3.11
37. R Core Team. R*: A Language and Environment for Statistical Computing, version 3.3. 1*. Vienna: R Foundation for Statistical Computing; 2016 (2019).
38. Frasso G, Lambert P. Bayesian inference in an extended SEIR model with nonparametric disease transmission rate: an application to the Ebola epidemic in Sierra Leone. *Biostatistics*. (2016) 17:779–92. doi: 10.1093/biostatistics/kxw027
39. EFSA. Guidance on good practice in conducting scientific assessments in animal health using modelling. *EFSA J*. (2009) 7:209–21. doi: 10.2903/j.efsa.2009.1419
40. Hudson EG, Brookes VJ, Ward MP. Demographic studies of owned dogs in the Northern Peninsula Area, Australia, to inform population and disease management strategies. *Austr Veterinary J*. (2018) 96:487–94. doi: 10.1111/avj.12766
41. Damman A, Viet A-F, Arnoux S, Guerrier-Chatellet M-C, Petit E, Ezanno P. Modelling the spread of bovine viral diarrhea virus (BVDV) in a beef cattle herd and its impact on herd productivity. *Vet Res*. (2015) 46:12. doi: 10.1186/s13567-015-0145-8
42. Tildesley MJ, Keeling MJ. Foot-and-mouth disease - a modelling comparison between the UK and Denmark. *Prevent Vetering Med*. (2008) 85:107–24. doi: 10.1016/j.prevetmed.2008.01.008
43. Kelso JK, Milne GJ. A spatial simulation model for the dispersal of the bluetongue vector Culicoides brevitarsis in Australia. *PLoS ONE*. (2014) 9:e104646. doi: 10.1371/journal.pone.0104646
44. Hudson EG, BrookesVJ, Dürr S, Ward MP. Modelling targeted rabies vaccination strategies for a domestic dog population with heterogeneous roaming patterns. *PLOS Negl Trop Dis*. (2019) 19:e0007582. doi: 10.1371/journal.pntd.0007582
45. Anderson RM, May RM. *Infectious Diseases of Humans*. New York, NY: Oxford Science Publications (1991).
46. Mur L, Sánchez-Vizcaíno JM, Fernández-Carrión E, Jurado C, Rolesu S, Feliziani F, et al. Understanding African swine fever infection dynamics in Sardinia using a spatially explicit transmission model in domestic pig farms. *Transbound Emerg. Dis*. (2018) 65:123–34. doi: 10.1111/tbed.12636
47. Kean JM, Barlow ND, Hickling GJ. Evaluating potential sources of bovine tuberculosis infection in a New Zealand cattle herd. *New Zealand J Agric Res*. (1999) 42:101–6. doi: 10.1080/00288233.1999.9513358
48. Vynnycky E, White R. *An Introduction to Infectious Disease Modelling*. Oxford University Press (2010).
49. Begon M, Bennett M, Bowers RG, French NP, Hazel SM, Turner J. A clarification of transmission terms in host-microparasite models: numbers, densities and areas. *Epidemiol Infect*. (2002) 129:147–53. doi: 10.1017/S0950268802007148
50. Ryder JJ, Miller MR, White A, Knell RJ, Boots M. Host-parasite population dynamics under combined frequency-and density-dependent transmission. *Oikos*. (2007) 116:2017–26. doi: 10.1111/j.2007.0030-1299.15863.x
51. Keeling MJ, Rohani P. *Modeling Infectious Diseases in Humans and Animals*. Princeton, NJ; Oxford: Princeton University Press, (2008). doi: 10.2307/j.ctvcm4gk0
52. McCallum H, Barlow N, Hone J. How should pathogen transmission be modeled? *Trends Ecol Evol*. (2001) 16:295–300. doi: 10.1016/S0169-5347(01)02144-9
53. Devenish-Nelson ES, Richards SA, Harris S, Soulsbury C, Stephens PA. Demonstrating frequency-dependent transmission of sarcoptic mange in red foxes. *Biol Lett*. (2014) 10:20140524. doi: 10.1098/rsbl.2014.0524
54. Kretzschmar M, Wallinga J. Mathematical models in infectious disease epidemiology. In: Mirjam Kretzschmar, editor. *Modern Infectious Disease Epidemiology*. New York, NY: Springer (2009). p. 209–21. doi: 10.1007/978-0-387-93835-6_12
55. Bates TW, Thurmond MC, Carpenter TE. Results of epidemic simulation modeling to evaluate strategies to control an outbreak of foot-and-mouth disease. *Am J Vet Res*. (2003) 64:205–10. doi: 10.2460/ajvr.2003.64.205
56. Ferguson NM, Donnelly CA, Anderson RM. The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science*. (2001) 292:1155–60. doi: 10.1126/science.1061020
57. Hirsch BT, Reynolds JJ, Gehrt SD, Craft ME. Which mechanisms drive seasonal rabies outbreaks in raccoons? A test using dynamic social network models. *J Appl Ecol*. (2016) 53:804–13. doi: 10.1111/1365-2664.12628
58. Sargent RG. Verification and validation of simulation models. *J Simulat*. (2013) 7:12–24. doi: 10.1057/jos.2012.20
59. Foddai A, Enøe C, Krogh K, Stockmarr A, Halasa T. Stochastic simulation modeling to determine time to detect bovine viral diarrhea

antibodies in bulk tank milk. *Prev Vet Med.* (2014) 117:149–59. doi: 10.1016/j.prevetmed.2014.07.007
60. Zinsstag J, Lechenne M, Laager M, Mindekem R, Naïssengar S, Oussiguéré A, et al. Vaccination of dogs in an African city interrupts rabies transmission and reduces human exposure. *Sci Transl Med*. (2017) 9:eaaf6984. doi: 10.1126/scitranslmed.aaf6984
61. Zingg D, Häsler S, Schuepbach-Regula G, Schwermer H, Dürr S. Evidence for emergency vaccination having played a crucial role to control the 1965/66 foot-and-mouth disease outbreak in Switzerland. *Front Vet Sci*. (2015) 2:72. doi: 10.3389/fvets.2015.00072
62. Halasa T, Bøtner A, Mortensen S, Christensen H, Toft N, Boklund A. Simulating the epidemiological and economic effects of an African swine fever epidemic in industrialized swine populations. *Vet Microbiol*. (2016) 193:7–16. doi: 10.1016/j.vetmic.2016.08.004
63. Cowled BD, Garner MG, Negus K, Ward MP. Controlling disease outbreaks in wildlife using limited culling: modelling classical swine fever incursions in wild pigs in Australia. *Vet Res.* (2012) 43:3. doi: 10.1186/1297-9716-43-3
64. Ezanno P, Fourichon C, Viet A-F, Seegers H. Sensitivity analysisto identify key-parameters in modelling the spread of bovine viraldiarrhoea virus in a dairy herd. *Prev Vet Med*. (2007) 80:49–64. doi: 10.1016/j.prevetmed.2007.01.005
65. Saltelli A. Making best use of model evaluations to compute sensitivity indices. *Computer Phys Commun*. (2002) 145:280–97. doi: 10.1016/S0010-4655(02)00280-1
66. Norton J. An introduction to sensitivity assessment of simulation models. *Environ Model Softw*. (2015) 69:166–74. doi: 10.1016/j.envsoft.2015.03.020
67. Frey CH, Patil SR. Identification and review of sensitivity analysis methods. *Risk Analysis*. (2002) 22:553–78. doi: 10.1111/0272-4332.00039
68. Grimm V, Berger U, DeAngelis DL, Polhill JG, Giske J, Railsback SF. The ODD protocol: a review and first update. *Ecol Model*. (2010) 221:2760–8. doi: 10.1016/j.ecolmodel.2010.08.019
69. Grimm V, Augusiak J, Focks A, Frank BM, Gabsi F, Johnston AS, et al. Towards better modelling and decision support: documenting model development, testing, and analysis using TRACE. *Ecol Model*. (2014) 280:129–39. doi: 10.1016/j.ecolmodel.2014.01.018
70. Mostert PF, Bokkers EAM, Van Middelaar CE, Hogeveen H, De Boer IJM. Estimating the economic impact of subclinical ketosis in dairy cattle using a dynamic stochastic simulation model. *Animal*. (2018) 12:145–54. doi: 10.1017/S1751731117001306
71. Webb CT, Carpenter TE, Dürr S, Ferrari M, Garner MG, Jewell C, et al. Ensemble modelling and structured decision-making to support emergency disease management. *Prev Vet Med*. (2017) 138:124–33. doi: 10.1016/j.prevetmed.2017.01.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Veterinary Science

REVIEW
published: 17 February 2021
doi: 10.3389/fvets.2020.539573

Sample Size Estimation in Veterinary Epidemiologic Research

*Mark A. Stevenson**

Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, VIC, Australia

In the design of intervention and observational epidemiological studies sample size calculations are used to provide estimates of the minimum number of observations that need to be made to ensure that the stated objectives of a study are met. Justification of the number of subjects enrolled into a study and details of the assumptions and methodologies used to derive sample size estimates are now a mandatory component of grant application processes by funding agencies. Studies with insufficient numbers of study subjects run the risk of failing to identify differences among treatment or exposure groups when differences do, in fact, exist. Selection of a number of study subjects greater than that actually required results in a wastage of time and resources. In contrast to human epidemiological research, individual study subjects in a veterinary setting are almost always aggregated into hierarchical groups and, for this reason, sample size estimates calculated using formulae that assume data independence are not appropriate. This paper provides an overview of the reasons researchers might need to calculate an appropriate sample size in veterinary epidemiology and a summary of sample size calculation methods. Two approaches are presented for dealing with lack of data independence when calculating sample sizes: (1) inflation of crude sample size estimates using a design effect; and (2) simulation-based methods. The advantage of simulation methods is that appropriate sample sizes can be estimated for complex study designs for which formula-based methods are not available. A description of the methodological approach for simulation is described and a worked example provided.

Keywords: sampling, epidemiiology, multilevel—hierarchical clustering, veterinary science, biostatistics

OPEN ACCESS

Edited by:
Heinzpeter Schwermer, Federal Food Safety and Veterinary Office (FSVO), Switzerland

Reviewed by:
Nikolaos Pandis, University of Bern, Switzerland
Ane Nødtvedt, Norwegian University of Life Sciences, Norway

***Correspondence:**
Mark A. Stevenson
mark.stevenson1@unimelb.edu.au

Specialty section:
This article was submitted to Veterinary Epidemiology and Economics, a section of the journal Frontiers in Veterinary Science

Received: *01 March 2020*
Accepted: *30 November 2020*
Published: *17 February 2021*

Citation:
Stevenson MA (2021) Sample Size Estimation in Veterinary Epidemiologic Research. Front. Vet. Sci. 7:539573. doi: 10.3389/fvets.2020.539573

INTRODUCTION

In the design of intervention and observational epidemiological studies sample size calculations are used to provide estimates of the minimum number of observations that need to be made to ensure that the stated objectives of a study are met (1, 2). Peer reviewed journals require investigators to provide justification of the number of subjects enrolled into a study and details of the assumptions and methodologies used to derive sample size estimates are now a mandatory component of grant application processes (3). Studies lacking in justification of sample size run the risk of failing to identify differences among treatment or exposure groups if a difference in those groups actually exist (4). Selection of a number of study subjects greater than that actually required results in a wastage of time and resources (2).

Methods for sample size estimation vary depending on the type of study being carried out i.e., observational (non-experimental) or interventional (experimental). Formula-based approaches for sample size estimation are often preferred by investigators because: (1) they

are relatively quick and simple to implement; (2) their widespread use makes peer review challenge less likely; and (3) the ability to use standard formulae goes hand in hand with "standard" study designs (i.e., randomized clinical trials, cross-sectional studies, case-control studies or cohort studies). Use of a standard study design implies the use of established approaches for data collection and analysis, again reducing the likelihood of challenge during peer review. In veterinary epidemiology the aggregation of animals into often several levels of hierarchy (e.g., cows within pens, pens within herds, herds within farms, and farms within regions) complicates sample size calculations due to lack of data independence arising from study subjects being aggregated into groups (e.g., pens, herds, farms, and regions). While modifications to standard sample size formulae are available, their flexibility to handle the range of real-world data situations is often limited.

The aim of this paper is to provide an overview of sample size estimation methods and their usage in applied veterinary epidemiological research. The structure of the paper is as follows. In the first section an overview of formula-based approaches for sample size estimation in epidemiological research is provided. In the second section, formula-based approaches for calculation of appropriate samples sizes for clustered data are presented. In the third and final section simulation-based approaches are presented as a means for estimating an appropriate sample size for hierarchical study designs for which formula-based methods are not available. Examples are provided throughout the paper to illustrate and support the concepts discussed. The supplementary material contains code allowing readers to reproduce the results presented in each of the examples using functions available in the contributed epiR package (5) in R (6).

FORMULA-BASED APPROACHES FOR SAMPLE SIZE ESTIMATION

In veterinary epidemiology sample size calculations are used during the design phase of a study to allow investigators to: (1) estimate a population parameter (e.g., the prevalence of disease); (2) test a hypothesis in an observational setting (using, for example, one of the three main observational study designs: cross-sectional, cohort or case-control); (3) test a hypothesis in an intervention setting (using a randomized clinical trial); and (4) achieve a specified level of confidence that an event will be detected if it is present at a specified design prevalence.

Sample Size Calculations to Estimate a Population Parameter

A summary of formula-based methods for estimation of a population parameter, all assuming data independence, is provided in **Tables 1**, **2**. Methods are defined for continuous and binary outcomes with different calculation methods dependent on the proposed sampling design: simple random, stratified

TABLE 1 | Information required to estimate a sample size for each of the common sampling designs, binary or continuous population parameters.

Outcome variable	Sampling design	Arguments	References
Continuous	Simple random	Total number of individual listing units in the population, the relative variance of the continuous variable to be estimated (i.e., the variance divided by the mean squared).	(7) pp. 74, Equation 3.14
Continuous	Stratified random	Total number of individual listing units in each strata, the expected means of the continuous variable to be estimated for each strata, the expected variances of the continuous variable to be estimated for each strata.	(7) pp. 176, Equation 6.25
Continuous	One stage cluster	Total number of clusters in the population, the population mean of the continuous variable to be estimated, the population variance of the continuous variable to be estimated.	(7) pp. 255, Box 9.4
Continuous	Two stage cluster	Number of individual listing units to be sampled from each cluster, the total number of clusters in the population and the number of individual listing units in each cluster, the mean of the continuous variable to be estimated at the first and second stage of sampling, the variance of the continuous variable to be estimated at the first and second stage of sampling.	(7) pp. 289, Equation 10.6
Binary	Simple random sampling	Total number of individual listing units in the population, the expected proportion of individual listing units with the outcome of interest.	(7) pp. 74, Equation 3.16
Binary	Stratified random	Total number of individual listing units in each strata, the expected proportion of individual listing units with the outcome of interest for each strata.	(7) pp. 176, Equation 6.23
Binary	One stage cluster	Total number of clusters in the population, the mean of the proportion of individual listing units in each cluster with the outcome of interest, the variance of the proportion of individual listing units in each cluster with the outcome of interest.	(7) pp. 255 Box 9.4
Binary	Two stage cluster	Number of individual listing units to be sampled from each cluster, the total number of clusters in the population and the number of individual listing units within each cluster, the mean of the denominator variable used to calculate the unknown population proportion at the first and second stage of sampling, the variance of the denominator variable used to calculate the unknown population proportion at the first and second stage of sampling, the variance of the numerator variable used to calculate the unknown population proportion at the first and second stage of sampling, the covariance of the unknown population proportion at the first and second stage of sampling.	(7) pp. 289, Equation 10.7

TABLE 2 | Formulae to estimate a sample size for each of the common sampling designs, binary or continuous population parameters.

Outcome variable	Sampling design	Formula	Arguments
Continuous	Simple random	$n \geq \frac{{z_{1-(\alpha/2)}}^2 \, N \, V_x^2}{{z_{1-(\alpha/2)}}^2 \, V_x^2 + (N-1) \, \epsilon_r^2}$	n = the number of subjects in the sample. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. N = the population size. V_x = the relative variance (the variance divided by the mean squared). ϵ_r = the relative error.
Continuous	Stratified random	$n \geq \frac{{z_{1-(\alpha/2)}}^2 \times \frac{N}{1+\gamma} \times V_x^2}{N\epsilon_r^2 + {z_{1-(\alpha/2)}}^2 \times \frac{V_x^2}{1+\gamma}}$	n = the number of subjects in the sample. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. N = the population size. γ = between strata variance σ_{bx}^2 divided by the within strata variance. σ_{wx}^2. V_x^2 = the relative variance (the variance divided by the mean squared). ϵ_r = the relative error.
Continuous	One stage cluster	$m = \frac{{z_{1-(\alpha/2)}}^2 M V_{1x}^2}{{z_{1-(\alpha/2)}}^2 V_{1x}^2 + (M-1)\epsilon_r^2}$ $V_{1x}^2 = \frac{\sigma_{1x}^2}{\bar{X}^2}$ $\sigma_{1x} = \frac{\sum_{i=1}^{M} (X_i - \bar{X})(Y_i - \bar{Y})}{M}$	m = the number of clusters in the sample $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. M = the number of clusters in the population. ϵ_r = the relative error. σ_{1x}^2 = the first stage variance components. $\bar{X}$ = mean level of X per cluster. X_i = level of the ith value of characteristic X.
Continuous	Two stage cluster	$m = \frac{\left(\frac{\sigma_{1x}^2}{\bar{X}^2}\right) \times \left(\frac{M}{M-1}\right) + \left(\frac{1}{\bar{n}}\right) \times \left(\frac{\sigma_{2x}^2}{\bar{X}^2}\right) \times \left(\frac{\bar{N}-\bar{n}}{\bar{N}-1}\right)}{\frac{\varepsilon_r^2}{{z_{1-(\alpha/2)}}^2} + \frac{\sigma_{1x}^2}{\bar{X}^2(M-1)}}$	m = the number of clusters in the sample. σ_{1x}^2 = the first stage variance components. $\bar{X}$ = mean level of X per cluster. M = the number of clusters in the population. $\bar{n}$ = the number of listing units to be sampled from each cluster. σ_{2x}^2 = the second stage variance components. $\bar{N}$ = the number of listing units in each cluster. ϵ_r = the relative error. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence.
Binary	Simple random sampling	$n \geq \frac{{z_{1-(\alpha/2)}}^2 \, N \, P_y \, (1-P_y)}{\left[(N-1)\,\epsilon_r^2 \, P_y^2\right] + {z_{1-(\alpha/2)}}^2 \, P_y(1-P_y)}$	n = the number of subjects in the sample. N = the population size P_y = the estimated population prevalence. ϵ_r = the relative error. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence.
Binary	Stratified random	$n \geq \frac{\left(\frac{{z_{1-(\alpha/2)}}^2}{N^2}\right) \sum_{h=1}^{L} \frac{N_h^2 \, P_{hy}(1-P_{hy})}{\pi_h P_y^2}}{\epsilon_r^2 + \left(\frac{{z_{1-(\alpha/2)}}^2}{N^2}\right)\left(\sum_{h=1}^{L} \frac{N_h P_{hy}(1-P_{hy})}{P_y^2}\right)}$ $\pi_h = \frac{n_h}{n}$	n = the number of subjects in the sample $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. N = the population size. L = the number of strata. N_h = the population size in the hth strata. P_{hy} = the estimated population prevalence in the hth strata. P_y = the estimated population prevalence. ϵ_r = the relative error. π_h = the fraction of samples allocated to strata h (decided in advance).
Binary	One stage cluster	When the number of listing units to be sampled per cluster is the same: $D = 1 + (b-1)\rho$ When the number of listing units to be sampled per cluster varies: $D = 1 + \left\{(CV^2+1)\,\bar{b} - 1\right\}\rho$ $n_c \geq \frac{{z_{1-(\alpha/2)}}^2 \, P_y \, (1-P_y) \, D}{(P_y \, \epsilon_r)^2 \bar{b}}$	D = the design effect. b = the number of listing units to be sampled from each cluster. ρ = the intracluster correlation coefficient. CV = the coefficient of variation of the number of listing units to be sampled from each cluster. $\bar{b}$ = the average number of listing units to be sampled from each cluster. n_c = the number of primary sampling units (clusters) to be sampled. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. P_y = the estimated population prevalence ϵ_r = the relative error.
Binary	Two stage cluster	$m = \frac{\left(\frac{\sigma_{1R}^2}{\bar{X}^2}\right) \times \left(\frac{M}{M-1}\right) + \left(\frac{1}{\bar{n}}\right) \times \left(\frac{\sigma_{2R}^2}{\bar{X}^2}\right) \times \left(\frac{\bar{N}-\bar{n}}{\bar{N}-1}\right)}{\frac{\varepsilon_r^2}{{z_{1-(\alpha/2)}}^2} + \frac{\sigma_{1R}^2}{\bar{X}^2(M-1)}}$	m = the number of clusters in the sample. σ_{1R}^2 = the first stage variance components. $\bar{X}$ = the mean level of characteristic X per listing unit. M = the number of clusters in the population. $\bar{n}$ = the number of listing units to be sampled from each cluster. σ_{2R}^2 = the first stage variance components. $\bar{X}$ = the mean level of characteristic X per cluster. $\bar{N}$ = the average number of listing units per cluster in the population. ϵ_r = the relative error. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence.

BOX 1 | The expected seroprevalence of brucellosis in a population of cattle is thought to be in the order of 15%. How many cattle need to be sampled and tested to be 95% certain that our seroprevalence estimate is within 20% (i.e., $0.20 \times 0.15 = 0.03$, 3%) of the true population value, assuming use of a test with perfect sensitivity and specificity? This formula requires the population size to be specified so we set N to a large number, 1,000,000:

$$n \geq \frac{z_{1-(\alpha/2)}{}^2 \, N P_y \, (1 - P_y)}{\left[(N-1)\, \epsilon_r^2 \, P_y^2\right] + z_{1-(\alpha/2)}{}^2 P_y \, (1 - P_y)}$$

$$n \geq \frac{1.96^2 \times 1,000,000 \times 0.15\,(1 - 0.15)}{\left[(1,000,000-1)\, 0.20^2 \, 0.15^2\right] + 1.96^2 \times 0.15 \times (1 - 0.15)}$$

$$n \geq \frac{489,804}{900.489}$$

$$n \geq 545$$

To be 95% confident that our estimate of brucellosis seroprevalence is within 20% of the true population value (i.e., a relative error of 0.20) 545 cattle should be sampled.

TABLE 3 | Information required to estimate a sample size for each of the common observational epidemiological study designs.

Study design	Arguments	References
Cross-sectional	The expected prevalence of the outcome among the exposed, the expected prevalence of the outcome among in the unexposed, the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects, sided test.	(9) pp. 313, Equation 8.14
Case-control	The expected odds ratio, the prevalence of exposure among controls, the required study power, the ratio of the number of control subjects to the number of case subjects, sided test.	(10)
Cohort, count data	The expected outcome incidence risk among the exposed, the expected outcome incidence risk among the unexposed, the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects, sided test.	(9) pp. 313, Equation 8.14
Cohort, time at risk	The expected outcome incidence rate among the exposed, the expected outcome incidence rate among the unexposed, the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects, sided test.	(11)

random, one-stage cluster and two-stage cluster designs. For continuous outcomes the analyst needs to provide an estimate of the mean of the outcome of interest and its expected variability. For binary outcomes only an estimate of the expected population proportion is required, given the variance of a proportion P equals $P \times [1 - P]$ (8). In addition to specifying the required level of confidence in the population parameter estimate (usually 95%) one needs to specify the desired maximum tolerable error. The maximum tolerable error is the difference between the true population parameter and the estimate of the true population parameter derived from sampling. In each of the formula-based approaches listed in **Tables 1**, **2** tolerable error is expressed in relative (as opposed to absolute) terms. If one assumes that the true population prevalence of disease is 0.40 and a desired relative tolerable error of 0.10 with 95% confidence is required, this means the calculation will return the required number of subjects to be 95% certain that the prevalence estimate from the study will be anywhere between $0.40 \pm (0.10 \times 0.40)$ that is, from 0.36 to 0.44. Some sample size formulae and/or software packages require maximum tolerable error to be expressed in absolute terms (that is, 0.04 for the example cited above). Analysts should take care to ensure that there is no ambiguity around the input format for tolerable error when using a published formula or software package since the distinction between absolute and relative error is often not clear in either the formula documentation or the graphic user interface, in the case of computer software. Similarly, when making a statement of the criteria used for sample size calculations when reporting the results of a study, care should be taken to ensure that the "relative" or "absolute" qualifier is used when referring to tolerable error.

In the absence of prior knowledge of the event prevalence in a population a conservative sample size estimate can be made assuming event prevalence is 0.5, since the variance of a prevalence (that is, $P \times [P - 1]$) is greatest when $P = 0.5$ and the absolute tolerable error and level of confidence remains fixed (8).

A worked example of a sample size calculation to estimate a prevalence using simple random sampling is shown in **Box 1**.

With stratified sampling the sampling frame is divided into groups (strata) and a random sample is taken from each stratum. When the variation of the outcome of interest within each stratum is small relative to the variation between strata, stratified random sampling returns a more precise estimate of the population parameter compared with simple random sampling.

Sample Size Calculations to Test a Hypothesis Using an Observational Study Design

Details of the formula-based methods to estimate a sample size for each of the main observational study (i.e., cross-sectional, case-control, and cohort studies) are provided in **Tables 3**, **4**. Again, these formulae all assume that data are independent.

TABLE 4 | Formulae to estimate a sample size for each of the common observational epidemiological study designs.

Study design	Formula	Arguments
Cross-sectional	$n \geq \frac{r+1}{r(\lambda-1)^2\pi^2}\left[z_{1-(\alpha/2)}\sqrt{(r+1)p_c(1-p_c)} + z_{1-\beta}\sqrt{\lambda\pi(1-\lambda\pi)+r\pi(1-\pi)}\right]^2$	n = the number of subjects in the sample. r = the anticipated number of subjects in the exposed group divided by the anticipated number of subjects in the unexposed group. λ = the expected prevalence ratio. π = the expected prevalence of the outcome among the non-exposed. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. p_c = the common prevalence over exposed and unexposed groups. $z_{1-\beta}$ = value from the standard normal curve corresponding to the desired study power. Use $z_{1-\beta} = -0.84$ for 80% power.
Case-control	$p_c^* = \frac{p_0}{r+1}\left(\frac{r\lambda}{1+(\lambda-1)p_0}+1\right)$ $n \geq \frac{(r+1)(1+(\lambda-1)p_0)^2}{rp_0^2(p_0-1)^2(\lambda-1)^2}\left[z_{1-(\alpha/2)}\sqrt{(r+1)p_c^*(1-p_c^*)} + z_{1-\beta}\sqrt{\frac{\lambda p_0(1-p_0)}{[1+(\lambda-1)p_0]^2}+rp_0(1-p_0)}\right]^2$	n = the number of subjects in the sample. p_0 = the expected prevalence of exposure among the controls. r = anticipated number of subjects in the control group divided by the anticipated number of subjects in the case group. λ = the expected odds ratio. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. $z_{1-\beta}$ = value from the standard normal curve corresponding to the desired study power. Use $z_{1-\beta} = 0.84$ for 80% power.
Cohort, count data	$n \geq \frac{r+1}{r(\lambda-1)^2\pi^2}\left[z_{1-(\alpha/2)}\sqrt{(r+1)p_c(1-p_c)} + z_{1-\beta}\sqrt{\lambda\pi(1-\lambda\pi)+r\pi(1-\pi)}\right]^2$	n = the number of subjects in the sample. r = the anticipated number of subjects in the exposed group divided by the anticipated number of subjects in the unexposed group. λ = the expected incidence risk ratio. π = the expected prevalence of the outcome among the non-exposed. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. p_c = the common prevalence over exposed and unexposed groups. $z_{1-\beta}$ = value from the standard normal curve corresponding to the desired study power. Use $z_{1-\beta} = 0.84$ for 80% power.
Cohort, time at risk	$\lambda_0' = \frac{\lambda_0^3 FT}{\lambda_0 FT - 1 + exp^{(-\lambda_0 FT)}}$ $\lambda_1' = \frac{\lambda_{10}^3 FT}{\lambda_1 FT - 1 + exp^{(-\lambda_1 FT)}}$ $\overline{\lambda}' = \frac{\overline{\lambda}^3 FT}{\overline{\lambda} FT - 1 + exp^{(-\overline{\lambda} FT)}}$ $n_A = m_B n_A \geq \frac{\left(z_{1-(\alpha/2)}\sqrt{(1+r)\overline{\lambda}'} + z_{1-\beta}\sqrt{(r \times \lambda_1' + \lambda_0')}\right)^2}{r(\lambda_1' - \lambda_0')^2}$	n_A = the number of subjects in the sample. λ_0 = the expected incidence rate among the unexposed. λ_1 = the expected incidence rate among the exposed. $\overline{\lambda} = (\lambda_0 + \lambda_1) / 2$ FT = the expected follow-up period for the study. r = anticipated number of subjects in the exposed group divided by the anticipated number of subjects in the unexposed group. $z_{1-(\alpha/2)}$ = value from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two-sided) confidence. $z_{1-\beta}$ = value from the standard normal curve corresponding to the desired study power. Use $z_{1-\beta} = 0.84$ for 80% power.

BOX 2 | A prospective cohort study of dry food diets and feline lower urinary tract disease (FLUTD) in mature male cats is planned. A sample of cats will be selected at random from the population and owners who agree to participate in the study will be asked to complete a questionnaire at the time of enrolment. Cats enrolled into the study will be followed for at least 5 years to identify incident cases of FLUTD. The investigators would like to be 0.80 certain of being able to detect when the risk ratio of FLUTD is 1.4 for cats habitually fed a dry food diet, using a 0.05 significance test. Previous evidence suggests that the incidence risk of FLUTD in cats not on a dry food (i.e., "other") diet is around 50 per 1000 per year. Assuming equal numbers of cats on dry food and other diets are sampled, how many cats should be enrolled into the study?

$$\lambda_0' = \frac{\lambda_0^3 FT}{\lambda_0 FT - 1 + exp^{(-\lambda_0 FT)}} \quad \lambda_1' = \frac{\lambda_1^3 FT}{\lambda_1 FT - 1 + exp^{(-\lambda_1 FT)}} \quad \bar{\lambda}' = \frac{\bar{\lambda}^3 FT}{\bar{\lambda} FT - 1 + exp^{(-\bar{\lambda} FT)}}$$

$$n_A \geq \frac{\left(z_{1-(\alpha/2)}\sqrt{(1+r)\,\bar{\lambda}'} + z_{1-\beta}\sqrt{(r \times \lambda_1' + \lambda_0')}\right)^2}{r(\lambda_1' - \lambda_0')^2}$$

$$n_A \geq \frac{\left(1.96\sqrt{(1+1)\,0.0264^2} + 0.84\sqrt{(1 \times 0.313 + 0.0217)}\right)^2}{1(0.07 - 0.05)^2}$$

$$n_A \geq \frac{(0.4509 + 0.1935)^2}{0.0004}$$

$$n_A \geq 1040$$

A total of 2,080 male cats need to be sampled to meet the requirements of the study (1,040 cats habitually fed dry food and 1,040 cats habitually fed "other" diet types).

BOX 3 | A case-control study of the association between white pigmentation around the eyes and ocular squamous cell carcinoma in Hereford cattle is planned. A sample of cattle with newly diagnosed squamous cell carcinoma will be compared for white pigmentation around the eyes with a sample of controls. Assuming an equal number of cases and controls, how many study subjects are required to detect an odds ratio of 2.0 with 0.80 power using a two-sided 0.05 test? Previous surveys have shown that around 0.30 of Hereford cattle without squamous cell carcinoma have white pigmentation around the eyes.

$$n \geq \frac{(r+1)(1+(\lambda-1)\,p_0)^2}{rp_0^2\,(p_0-1)^2\,(\lambda-1)^2}\left[z_{1-(\alpha/2)}\sqrt{(r+1)p_C^*(1-p_C^*)} + z_\beta\sqrt{\frac{\lambda p_0(1-p_0)}{[1+(\lambda-1)\,p_0]^2} + rp_0(1-p_0)}\right]^2$$

$$n \geq \frac{(1+1)(1+(2-1)\,0.3)^2}{1 \times 0.3^2\,(0.3-1)^2\,(2-1)^2}\left[1.96\sqrt{(1+1)\,0.38(1-0.38)} + 0.84\sqrt{\frac{2 \times 0.3(1-0.3)}{[1+(2-1)\,0.3]^2} + 1 \times 0.3(1-0.3)}\right]^2$$

$$n \geq \frac{3.38}{0.0441}\,[1.346 + 0.569]^2$$

$$n \geq 282$$

If the true odds for squamous cell carcinoma in exposed subjects relative to unexposed subjects is 2.0, we will need to enrol 141 cases and 141 controls (282 cattle in total) to reject the null hypothesis that the odds ratio equals one with probability (power) 0.80. The Type I error probability associated with this test of this null hypothesis is 0.05.

Note that the sample size formulae for cross-sectional studies, cohort studies using count data and cohort studies using time at risk require the analyst to provide an estimate of prevalence, incidence risk and incidence rate (respectively) for both risk factor exposed and unexposed groups. **Box 2** provides a worked example for a prospective cohort study, with a fixed follow-up time.

In contrast to sample size formulae for cross-sectional and cohort studies, the sample size formula for case-control studies requires provision of an estimate of the prevalence of exposure amongst controls (**Box 3**). An additional consideration when estimating an appropriate sample size for a case-control study is specification of the design – either matched or unmatched (12). The process of matching provides a means for controlling for the effect of a known confounder with the added benefit of an increase in statistical efficiency (12, 13).

Sample Size Calculations to Test a Hypothesis Using a Randomized Clinical Trial

A superiority trial is a study in which the aim is to show that a treatment intervention provides a better therapeutic outcome than a known reference (often a placebo) and the statistical procedure to provide this evidence is called a superiority test (14).

In situations where an established treatment already exists a study comparing a new treatment to a placebo (effectively, no treatment) will be unethical. In this situation interest lies

TABLE 5 | Information required to estimate a sample size for equivalence, superiority and non-inferiority trials.

Outcome variable	Study design	Arguments	References
Continuous	Equivalence trial	The expected mean of the outcome variable in the treatment and control groups, the expected population standard deviation of the outcome variable, the equivalence limit (expressed as a proportion), the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects.	(15–17).
Binary	Equivalence trial	The expected proportion of successes in the treatment and control groups, the equivalence limit (expressed as a proportion), the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects.	(15–17).
Continuous	Superiority trial	The expected mean of the outcome variable in the treatment and control groups, the expected population standard deviation of the outcome variable, the equivalence limit (expressed as a proportion), the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects.	(15).
Binary	Superiority trial	The expected proportion of successes in the treatment and control groups, the equivalence limit (expressed as a proportion), the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects.	(15)
Continuous	Non-inferiority trial	The expected mean of the outcome variable in the treatment and control groups, the expected population standard deviation of the outcome variable, the equivalence limit (expressed as a proportion), the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects.	(15–17)
Binary	Non-inferiority trial	The expected proportion of successes in the treatment and control groups, the equivalence limit (expressed as a proportion), the required study power, the ratio of the number of exposed subjects to the number of unexposed subjects.	(16, 17)

TABLE 6 | Formulae to estimate a sample size for equivalence, superiority and non-inferiority trials, binary or continuous population parameters.

Outcome variable	Study design	Formula	Arguments
Continuous	Equivalence trial	$n_A = rn_B$ $n_B = \left(1 + \frac{1}{r}\right)\left(\sigma \frac{z_{1-(\alpha/2)} + z_{1-\beta/2}}{\lvert\mu_A - \mu_B\rvert - \delta}\right)^2$ $1 - \beta = 2\left[\Phi\left(z - z_{1-(\alpha/2)}\right) + \Phi\left(-z - z_{1-(\alpha/2)}\right)\right] - 1$ $z = \frac{\lvert\mu_A - \mu_B\rvert - \delta}{\sigma\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$	μ_A = the expected mean of the outcome in the treatment group. μ_B = the expected mean of the outcome in the control group. σ = the expected standard deviation of the outcome across treatment and control groups. r = anticipated number of subjects in the treatment group divided by the anticipated number of subjects in the control group. Φ = the standard Normal distribution function. Φ^{-1} = the standard Normal quantile function. α = the Type I error, e.g. $\alpha = 0.05$. β = the Type II error, e.g. $\beta = 0.20$. δ = the equivalence margin.
Binary	Equivalence trial	$n_A = rn_B$ $n_B = \left(\frac{p_A(1-p_A) + p_B(1-p_B)}{r}\right)\left(\sigma \frac{z_{1-(\alpha/2)} + z_{1-\beta/2}}{\lvert p_A - p_B\rvert - \delta}\right)^2$ $1 - \beta = 2\left[\Phi\left(z - z_{1-(\alpha/2)}\right) + \Phi\left(-z - z_{1-(\alpha/2)}\right)\right] - 1$ $z = \frac{\lvert p_A - p_B\rvert - \delta}{\sigma\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}}$	p_A = the expected probability of success in the treatment group. p_B = the expected probability of success in the control group. r = anticipated number of subjects in the treatment group divided by the anticipated number of subjects in the control group. Φ = the standard Normal distribution function. Φ^{-1} = the standard Normal quantile function. α = the Type I error, e.g. $\alpha = 0.05$. β = the Type II error, e.g. $\beta = 0.20$. δ = the equivalence margin.
Continuous	Superiority trial or non-inferiority trial	$n_A = rn_B$ $n_B = \left(1 + \frac{1}{r}\right)\left(\sigma \frac{z_{1-(\alpha/2)} + z_{1-\beta}}{\mu_A - \mu_B - \delta}\right)^2$ $1 - \beta = \Phi(z - z_{1-\alpha}) + \Phi(-z - z_{1-\alpha})$ $z = \frac{\mu_A - \mu_B - \delta}{\sigma\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$	μ_A = the expected mean of the outcome in the treatment group. μ_B = the expected mean of the outcome in the control group. σ = the expected standard deviation of the outcome across treatment and control groups. r = anticipated number of subjects in the treatment group divided by the anticipated number of subjects in the control group. Φ = the standard Normal distribution function. Φ^{-1} = the standard Normal quantile function. α = the Type I error, e.g. $\alpha = 0.05$. β = the Type II error, e.g. $\beta = 0.20$. δ = the equivalence margin.
Binary	Superiority trial or non-inferiority trial	$n_A = rn_B$ $n_B = \left(\frac{p_A(1-p_A) + p_B(1-p_B)}{r}\right)\left(\frac{z_{1-(\alpha/2)} + z_{1-\beta}}{p_A - p_B - \delta}\right)^2$ $1 - \beta = \Phi\left(z - z_{1-(\alpha/2)}\right) + \Phi\left(-z - z_{1-(\alpha/2)}\right)$ $z = \frac{p_A - p_B - \delta}{\sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}}$	p_A = the expected probability of success in the treatment group. p_B = the expected probability of success in the control group. r = anticipated number of subjects in the treatment group divided by the anticipated number of subjects in the control group. Φ = the standard Normal distribution function. Φ^{-1} = the standard Normal quantile function. α = the Type I error, e.g. $\alpha = 0.05$. β = the Type II error, e.g. $\beta = 0.20$. δ = the equivalence margin.

$z_{1-(\alpha/2)}$ = values from the standard normal curve corresponding to the desired level of confidence. Use $z_{1-(\alpha/2)} = 1.96$ for 95% (two side) confidence.

$z_{1-\beta}$ = value from the standard normal curve corresponding to the desired study power. Use $z_{1-\beta} = 0.84$ for 80% power.

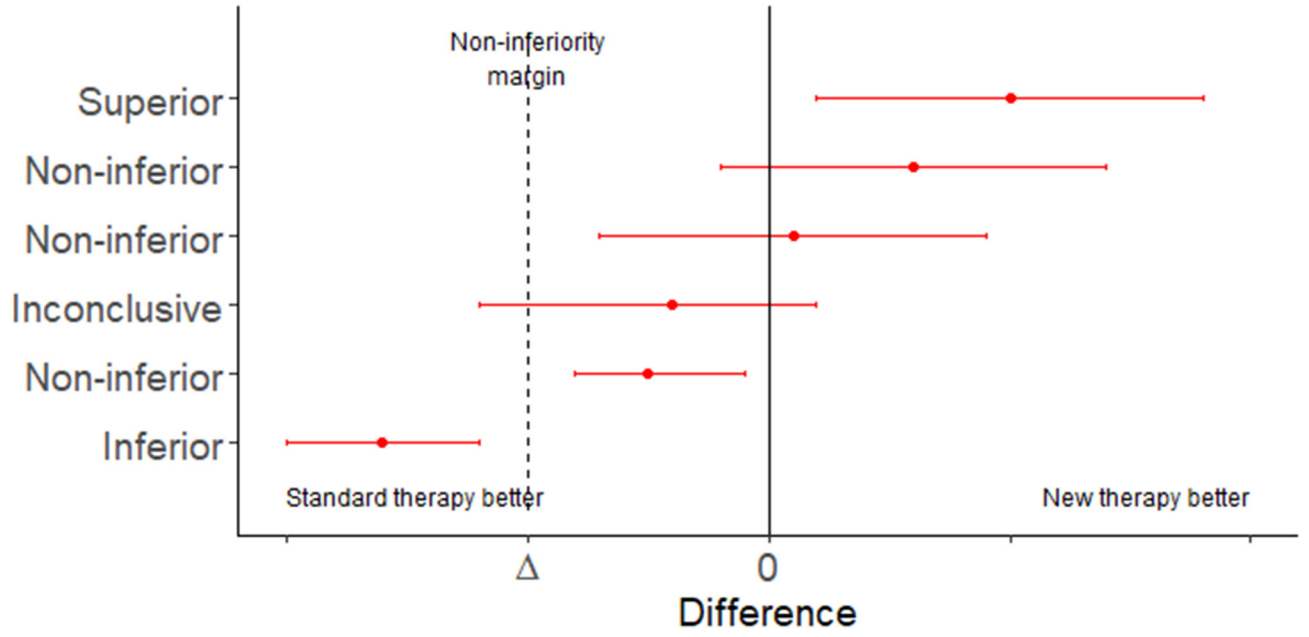


FIGURE 1 | Error bar plot showing the possible conclusions to be drawn from a non-inferiority trial. Adapted from Head et al. (18).

BOX 4 | Suppose a pharmaceutical company would like to conduct a clinical trial to compare the efficacy of two antimicrobial agents when administered orally to patients with skin infections. Assume the true mean cure rate of the treatment is 0.85 and the true mean cure rate of the control is 0.65. We consider a difference of <0.10 in cure rate to be of no clinical importance (i.e., delta = −0.10). Assuming a one-sided test size of 5% and a power of 80% how many subjects should be included in the trial?

$$n_B = \left(\frac{p_A(1-p_A)+p_B(1-p_B)}{r}\right)\left(\frac{z_{1-(\alpha/2)}+z_{1-\beta}}{p_A-p_B-\delta}\right)^2$$

$$n_B = \left(\frac{0.85(1-0.85)+0.65(1-0.65)}{1}\right)\left(\frac{1.96+0.84}{0.85-0.65--0.10}\right)^2$$

$$n_B = \left(\frac{0.355}{1}\right)\left(\frac{2.48}{0.30}\right)^2$$

$$n_B = 25$$

A total of 50 subjects need to be enrolled in the trial, 25 in the treatment group and 25 in the control group.

in determining if the new treatment is: (1) either the same as, or better than, an established treatment using a non-inferiority trial; or (2) equivalent to an existing treatment within a specified range, using an equivalence trial (**Tables 5**, **6** and **Figure 1**). Equivalence trails are not to be confused with bioequivalence trials where generic drug preparations are compared to currently marketed formulations with respect to their pharmacokinetic parameters.

Superiority, non-inferiority and equivalence trials require the analyst to specify an equivalence margin. The equivalence margin is the range of values for which the treatment efficacies are close enough to be considered the same (19). Expressed in another way, the equivalence margin is the maximum clinically acceptable difference one is willing to accept in return for the secondary benefits of the new treatment.

Equivalence margins can be set on the basis of a clinical estimation of a minimally important effect. This approach is subjective and, as a result, it is possible to set the equivalence margin to be greater than the effect of the established treatment, which could lead to potentially harmful treatments classified as non-inferior. A second approach is to select an equivalence margin with reference to the effect of the established treatment in trials where a placebo has been used. When the equivalence margin is chosen in this way, there is some objective basis on which to claim that a positive non-inferiority trial implies that a new treatment is, in fact, superior to the established treatment (assuming the effect of the established treatment in the current

TABLE 7 | Summary of sample size formulae to estimate the probability of disease freedom or estimate surveillance system sensitivity.

Study design	Outcome	Arguments	References
Representative sampling	Probability of disease freedom assuming imperfect test sensitivity and perfect test specificity.	The assumed population size, an estimate of the prior probability that the population is free of disease, an estimate of the probability of disease introduction, the population level design prevalence, the desired probability that the population is free of disease, the sensitivity of the diagnostic test used at the surveillance unit level.	(22, 23)
Representative sampling	Surveillance system sensitivity assuming a single risk factor and varying test sensitivity.	The assumed population size, the population level design prevalence, the desired surveillance system sensitivity, the sensitivity of the diagnostic test used at the surveillance unit level.	(24, 25)
Representative sampling	Surveillance system sensitivity assuming two stage sampling, imperfect test sensitivity and perfect test specificity.	The number of clusters in the population, the number of surveillance units within each cluster, the cluster level design prevalence, the desired cluster level sensitivity, the population level design prevalence, the desired population level sensitivity, the sensitivity of the diagnostic test used at the surveillance unit level.	(24–27)
Representative sampling	Surveillance system sensitivity, imperfect test sensitivity and imperfect test specificity.	The assumed population size, the population level design prevalence, the desired population level sensitivity, the desired population level specificity, the sensitivity of the diagnostic test at the surveillance unit level, the specificity of the diagnostic test at the surveillance unit level.	(26–28)
Representative sampling	Surveillance system sensitivity assuming pooled sampling giving rise to imperfect test sensitivity and imperfect test specificity.	The number of surveillance units that contribute to each pool, the population level design prevalence, the sensitivity of the diagnostic test at the pooled level, the specificity of the diagnostic test at the pooled level, the desired population level sensitivity.	(29)
Risk-based sampling	Surveillance system sensitivity assuming imperfect test sensitivity and perfect test specificity.	The population level design prevalence, relative risk estimates for each strata, the population proportions for each strata, the surveillance proportions for each strata, the desired population level sensitivity, the sensitivity of the diagnostic test at the surveillance unit level.	(5)
Risk-based sampling	Surveillance system sensitivity assuming risk-based 2-stage sampling on one risk factor at the cluster level assuming imperfect test sensitivity and perfect test specificity.	Relative risk values for each strata in the population, the population proportions in each strata, the planned number of units to be sampled from each strata, the cluster level design prevalence, the desired cluster level sensitivity, the surveillance unit level design prevalence, the sensitivity of the diagnostic test at the surveillance unit level, the desired surveillance system (population-level) sensitivity.	(5)
Risk-based sampling	Surveillance system sensitivity assuming risk-based 2-stage sampling on two risk factors at either the cluster level, the unit level or both, imperfect test sensitivity and perfect test specificity.	The number of risk strata defining the relative risk values at the cluster level, the population proportions at the cluster level, the planned surveillance proportions at the cluster level, the cluster level design prevalence, the desired cluster level sensitivity, the number of risk strata defining the relative risk values at the surveillance unit level, the population proportions at the surveillance unit level, the planned surveillance proportions at the surveillance unit level, the surveillance unit level design prevalence, the sensitivity of the diagnostic test at the surveillance unit level, the desired surveillance system (population-level) sensitivity.	(5)

trial is similar to its effect in the historical trials). An example sample size calculation for a non-inferiority trial is presented in **Box 4**.

Sample Size Calculations to Detect the Presence of an Event

Sampling of individuals to either detect the presence of an event (usually the presence of disease or the presence of infection) or provide evidence that disease is absent from a jurisdiction are frequent activities in veterinary epidemiology. Typical scenarios include: (1) shipment of live animals from one country to another where the country receiving the shipment might request that testing is carried out on a sample of individuals, as opposed to testing every animal; and (2) a country wishing to re-gain official disease freedom status following an infectious disease outbreak.

Details of formula-based sample size estimation methods to detect the presence of an event are provided in **Table 7**. Sample size estimation methods can be categorized into two groups: (1) to ensure sufficient units are sampled to return a desired (posterior) probability of disease freedom; and (2) to ensure sufficient units are sampled to ensure a surveillance system has a desired system sensitivity. For the surveillance system sensitivity methods sampling can be either representative or risk based. All of the methods listed in **Table 7** account for imperfect diagnostic test sensitivity at the surveillance unit level.

When tests with both imperfect diagnostic sensitivity and specificity are used, diseased individuals can be missed because of imperfect diagnostic sensitivity but at the same time disease negative individuals can be incorrectly identified as disease positive because of imperfect specificity. Cameron and Baldock (26) describe an approach to estimate the number of animals to be sampled from a finite population using a test with imperfect diagnostic sensitivity and specificity using the hypergeometric distribution. This method returns the number of individuals to be sampled and the estimated probability that the population is diseased for 1 to n individuals that return a positive test result. This allows an analyst to make a statement that they can be (for example) 95% confident that the prevalence of disease in the population of interest is less than the stated design prevalence if the number of (surveillance) units with a positive test result is less than or equal to a specified cut-point. A worked example of this approach is provided in the **Supplementary Material**.

BOX 5 | An aid project has distributed cook stoves in a single province in a resource-poor country. At the end of 3 years, the donors would like to know what proportion of households are still using their donated stove. A cross-sectional study is planned where villages in a province will be sampled and all households (∼75 per village) will be visited to determine if the donated stove is still in use. A pilot study of the prevalence of stove usage in five villages showed that 0.46 of householders were still using their stove and the ICC for stove use within villages is in the order of 0.20. If the donor wanted to be 95% confident that the survey estimate of stove usage was within 10% of the true population value, how many villages (clusters) need to be sampled?

$$D = 1 + (b - 1)\rho$$
$$D = 1 + (75 - 1) \times 0.20$$
$$D = 15.8$$
$$n_c \geq \frac{z_{1-(\alpha/2)}^2 P_y (1 - P_y) D}{(P_y \epsilon_r)^2 b}$$
$$n_c \geq \frac{1.96^2\ 0.46\ (1 - 0.46)\ 15.8}{(0.46 \times 0.10)^2 \times 75}$$
$$n_c \geq \frac{15.077}{0.1587}$$
$$n_c \geq 96$$

A total of 96 villages need to be sampled to meet the requirements of the study.

BOX 6 | Continuing the example presented in Box 5, we are now told that the number of households per village varies. The average number of households per village is 75 with a 0.025 quartile of 40 households and a 0.975 quartile of 180. Assuming the number of households per village follows a normal distribution the expected standard deviation of the number of households per village is in the order of (180 – 40) ÷ 4 = 35. How many villages need to be sampled? In the formula below, CV standards for coefficient of variation defined as the standard deviation of the cluster sizes divided by the mean of the cluster sizes.

$$D = 1 + \left\{ (CV^2 + 1)\ b - 1 \right\} \rho$$
$$D = 1 + \left\{ \left(0.467^2 + 1 \right)\ 75 - 1 \right\} 0.2$$
$$D = 19.1$$
$$n_c \geq \frac{z_{1-(\alpha/2)}^2 P_y (1 - P_y) D}{(P_y \epsilon_r)^2 b}$$
$$n_c \geq \frac{1.96^2\ 0.46\ (1 - 0.46)\ 19.1}{(0.46 \times 0.10)^2 \times 75}$$
$$n_c \geq \frac{18.194}{0.1587}$$
$$n_c \geq 115$$

A total of 115 villages need to be sampled to meet the requirements of the study.

BOX 7 | Continuing the example provided in Box 1, being seropositive to brucellosis is likely to cluster within herds. Otte and Gumm (20) cite the intracluster correlation coefficient for *Brucella abortus* in cattle to be in the order of 0.09. We now adjust our sample size estimate of 545 to account for clustering at the herd level. Assume that, on average, $b = 20$ animals will be sampled per herd:

$$D = 1 + (b - 1)\rho$$
$$D = 1 + (20 - 1) \times 0.09$$
$$D = 2.71$$

After accounting for the presence of clustering at the herd level we estimate that a total of (545 × 2.71) = 1,477 cattle need to be sampled to meet the requirements of the survey. If 20 cows are sampled per herd this means that a total of (1,477 ÷ 20) = 74 herds are required.

BOX 8 | Dohoo et al. (21) provide details of an observational study of the reproductive performance of dairy cows on Reunion Island. If this study were to be repeated, how many lactations would need to be sampled to be 95% confident that the estimated logarithm of calving to conception interval was within 5% of the true population value?

From (21) the standard deviations of the random effect terms from a multilevel model of factors influencing log transformed calving to conception interval at the herd, cow and lactation level were 0.1157, 0.1479, and 0.5116, respectively. The ICC for lactations within herds (Equation 3):

$$\rho_2 = \frac{\sigma_3^2}{\sigma_3^2+\sigma_2^2+\sigma_1^2}$$
$$\rho_2 = \frac{0.1157^2}{0.1157^2 + 0.1479^2 + 0.5116^2}$$
$$\rho_2 = 0.0451$$

and the ICC for lactations within cows (Equation 4):

$$\rho_1 = \frac{\sigma_3^2+\sigma_2^2}{\sigma_3^2+\sigma_2^2+\sigma_1^2}$$
$$\rho_1 = \frac{0.1157^2 + 0.1479^2}{0.1157^2 + 0.1479^2 + 0.5116^2}$$
$$\rho_1 = 0.1188$$

The mean and standard deviation of the logarithm of calving to conception interval was 4.59 and 0.54, respectively. What is the required sample size assuming the data are independent?

$$m_1 \geq \frac{1.96 \times 1,000,000 \times (0.54^2 \div 4.59^2)}{1.96\,(0.54^2 \div 4.59^2) + ([1,000,000 - 1]\ 0.05^2)}$$
$$m_1 \geq \frac{54057.4}{2500.025}$$
$$m_1 \geq 22$$

Assuming the data are independent a total of 22 lactations are required to be 95% confident that our estimate of the logarithm of calving to conception interval is within 5% of the true population value.

We elect to sample two lactations per cow. How many lactations are required to account for clustering of lactations within cows?

$$n_1 = 2$$
$$D_1 = 1 + \rho_1\,(n_1 - 1)$$
$$D_1 = 1 + 0.1188\,(2 - 1)$$
$$D_1 = 1.1188$$
$$m_2 = D_1 \times m_1$$
$$m_2 = 1.1188 \times 22$$
$$m_2 = 25$$

A total of 25 lactations are required accounting for clustering of lactations within cows. How many cows are required?

$$n_2 = m_2 / n_1$$
$$n_2 = 25 / 2$$
$$n_2 = 13$$

A total of 13 cows are required if we sample two lactations per cow (26 lactations in total).

We now consider clustering at the herd level. How many lactations are required to account for clustering of cows within herds?

$$D_2 = 1 + (n_1 \times (n_2 - 1) \times \rho_2) + ((n_1 - 1) \times \rho_1)$$
$$D_2 = 1 + (2 \times (13 - 1) \times 0.0451) + ((2 - 1) \times 0.1188)$$
$$D_2 = 2.2016$$
$$m_3 = D_2 \times m_1$$
$$m_3 = 2.2061 \times 22$$
$$m_3 = 49$$

Continued

BOX 8 | **Continued**

Accounting for clustering of lactations within cow and cows within herds, a total of 39 lactations are required. How many herds are required?

$$n_3 = m_3 / (n_1 \times n_2)$$
$$n_3 = 49 / (2 \times 13)$$
$$n_3 = 2$$

A total of 2 herds are required if we sample 13 cows from each herd and 2 lactations from each cow. The total number of lactations required is therefore:

$$n_{total} = (n_1 \times n_2 \times n_3)$$
$$n_{total} = (2 \times 13 \times 2)$$
$$n_{total} = 52$$

To account for lack of independence in the data arising from clustering of lactations within cows and cows within herds 52 lactations (2 lactations from 13 cows from 2 herds) are required to meet the requirements of the study.

The required sample size assuming the data were independent was 22. The required sample size accounting for lack of independence in the data was 52, a 2.5-fold difference.

SAMPLE SIZE CALCULATIONS FOR CLUSTERED DATA

Aggregation of individual sampling units into groups ("clusters") for example farms, households or villages violates the assumption of independence that is central to the sample size calculation methods described so far. When individuals are aggregated into clusters there are two sources of variation in the outcome of interest. The first arises from the effect of the cluster; the second from the effect of the individual. This means that individuals selected from the same cluster are more likely to be similar compared with those sampled from the general population (30). For this reason, the effective sample size when observations are made on randomly selected individuals from the same cluster will be less than that when observations are made on individuals selected completely at random from the general population. For studies where the objective is to estimate a population parameter (e.g., a prevalence) a reduction in effective sample size increases the uncertainty around the estimate of the population parameter. For studies where the objective is to test a hypothesis a reduction in effective sample size results in a reduction in statistical power, in effect the ability to detect a statistically significant difference in event outcomes for exposure positive and exposure negative individuals given a true difference actually exists.

With one-stage cluster sampling a random sample of clusters is selected first and then all individual listing units within each cluster are selected for study. With two-stage cluster sampling a random sample of clusters is selected first and then a random sample of individual listing units within each cluster is selected. The primary advantage of cluster sampling is logistics. In animal health, where animals are typically managed within clusters (e.g., herds or flocks) it is easier to select clusters first and then, from each selected cluster, take a sample of individual animals. This contrasts with a simple random sampling approach which would require an investigator to travel to a large number of herds-flocks, sampling small numbers of animals from each. As explained above, the main disadvantage of cluster sampling is a reduction in the effective sample size due to animals from the same cluster being more homogenous (similar) compared with those from different clusters.

To compensate for this lack of precision Donner et al. (31) proposed that a sample size estimate calculated assuming complete independence (using the formulae presented in **Tables 1**–**7**) can be inflated by a value known as the design effect (D) to achieve the level of statistical power achieved using independent sampling. For a single level of clustering (e.g., the situation where cows are clustered within herds) the design effect is calculated as:

$$D = 1 + (b - 1)\rho \quad (1)$$

In Equation 1 b equals the number of animals to be sampled from each cluster (not to be confused with the total number of animals eligible for sampling within each cluster) and ρ is the intracluster correlation coefficient (ICC). The value of ρ equals the between-cluster variance σ_B^2 divided by the between-cluster variance plus the within-cluster variance $(\sigma_B^2 + \sigma_W^2)$:

$$\rho = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} \quad (2)$$

When there is little variation in an outcome within a cluster (e.g., observations made on individual cows within herds are "similar," σ_W^2 will be small) ρ will be close to 1 and the design effect will therefore be large. When there is wide variation within a cluster (e.g., observations made on individual cows within herds showing a similar variability to the general population, σ_W^2 will be large) ρ will be close to 0 and, therefore, the design effect will be close to unity. Using the definition provided above (Equation 2), ρ ranges between 0 and +1 with typical values ranging from 0 to 0.05 for non-communicable diseases and values >0.4 uncommon. Papers providing ICC estimates for various outcomes in the human

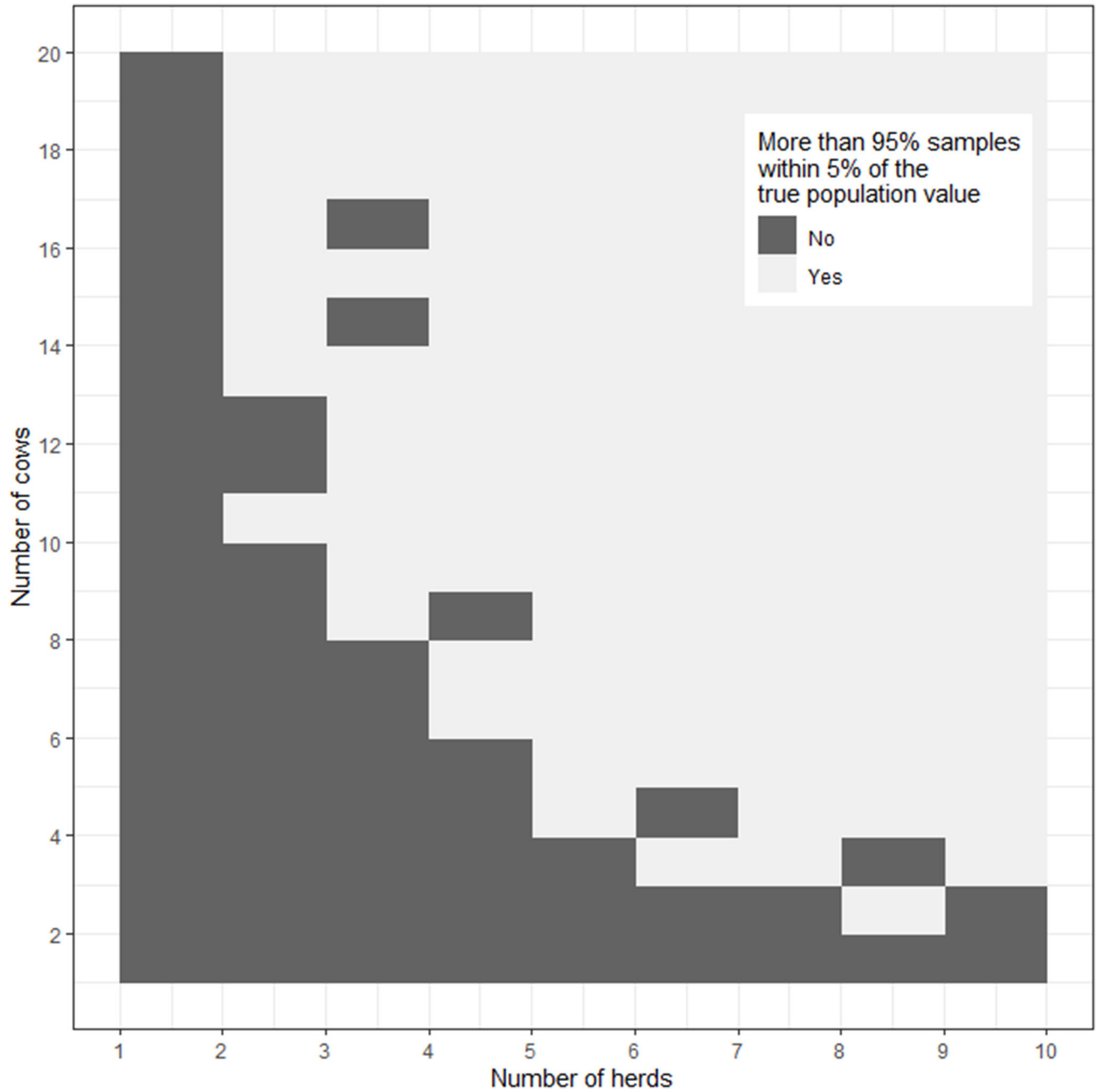


FIGURE 2 | Image plot showing the percentage of simulations where the sample estimate of the logarithm of calving to conception interval was within 5% of the true population value as a function of the number of herds and number of cows from each herd sampled. The dashed line shows the herd-cow sample size combinations where >95% of simulations returned an estimate of the logarithm of calving to conception interval that was within 5% of the true population value.

and veterinary literature have been published: see, for example, (20, 32, 33). Researchers should be aware of the importance of publishing estimates of ICC since high quality empirical data are necessary to provide credible sample size estimates for future studies. More importantly, for the same outcome measure, ICC estimates will vary from one research setting to another so access to a likely range of ICC measures is desirable.

A number of methods are available to estimate ρ from empirical data (34, 35) ranging from one-way analysis of variance (36) to regression-based approaches using mixed effects models (37). Eldridge and Kerry (38) provide a comprehensive review of appropriate techniques.

An example of how the ICC can be used to estimate the number of primary sampling units for a one-stage cluster design is provided in **Box 5**.

The example shown in **Box 5** is somewhat unrealistic in that it is assumed that the number of households in each village is a constant value of 75. Eldridge, Ashby and Kerry (39) provide an approach to estimate a sample size using a one-stage cluster design when the number of individual listing units per cluster varies (**Box 6**).

An example showing how a crude sample size estimate (i.e., a sample size calculated assuming independence) can be adjusted to account for clustering using the design effect is provided in **Box 7**.

Three levels of clustering are relatively common in veterinary epidemiological research (much more so than in human epidemiology) where, for example, lactations (level 1 units) might be sampled within cows (level 2 units) which are then sampled within herds (level 3 units). The total variance in this situation is made up of the variance associated with lactations within cows within herds σ_1^2, the variance between cows within herds σ_2^2, and the variance between herds σ_3^2. Two ICCs can be calculated: lactations within herds:

$$\rho_2 = \frac{\sigma_3^2}{\sigma_3^2 + \sigma_2^2 + \sigma_1^2} \tag{3}$$

and lactations within cows:

$$\rho_1 = \frac{\sigma_3^2 + \sigma_2^2}{\sigma_3^2 + \sigma_2^2 + \sigma_1^2} \tag{4}$$

In a study comprised of three levels the required sample size, accounting for clustering equals (40):

$$n_3 n_2 n_1 = DE \times m \tag{5}$$

In Equation 5, m is the number of lactations to be sampled to meet the requirements of the study assuming the data are completely independent and n_3, n_2, and n_1 are the number of units to be sampled at the herd, cow and lactation level (respectively). The design effect for three levels of clustering equals:

$$DE = 1 + n_1 (n_2 - 1) \rho_2 + (n_1 - 1) \rho_1 \tag{6}$$

Box 8 provides a worked example of a sample size calculation for the three-level clustering scenario.

SIMULATION-BASED APPROACHES FOR SAMPLE SIZE ESTIMATION

In applied veterinary epidemiological research it is common for study designs not to conform to the standard study designs for which sample size formulae are available. Typical examples include situations where study subjects are organized into more than three levels of aggregation and in clinical trials where a treatment might be applied at the group level and a second treatment applied at the individual level. Where there are multiple levels of aggregation researchers may elect to apply a more conservative design effect multiplier than that used when study subjects are kept in simpler cluster groups. While this approach is an attempt to address the problem, it can result in the final sample size estimate being larger than the final sample size required if the design effect was known more precisely.

For complex study designs simulation-based approaches provide an alternative for sample size estimation that is relatively easy to implement using modern statistical software (41). In the text that follows a worked example is provided, where simulation is used to estimate the number of lactations, cows and herds to be sampled to provide an estimate of log calving to conception interval, using the scenario presented in **Box 8**.

The general approach when using simulation to estimate a sample size to estimate a population parameter is to: (1) simulate a population data set that respects clustering of the outcome variable within the population of interest; (2) define a series of candidate sample size estimates; (3) repeatedly sample the simulated population using each of the candidate sample size estimates to determine the proportion of occasions the estimate of the population parameter is within the prescribed relative error of the true population value. When estimating a population prevalence and assuming the level of confidence specified by the analyst has been set to 95%, the required sample size is the combination of level 1, 2, 3, ... n units sampled that returns an estimate of the outcome variable that is within the prescribed relative error of the true population value on 95% of occasions. Note that several different combinations of units sampled at each level might achieve the stated objectives of the study.

When the study aim is to test a hypothesis, an additional step is to assign the exposure variable (e.g., a treatment) to members of the population and then to estimate the effect of the exposure on the outcome of interest using a regression approach. Arnold et al. (41) provide a worked example of this approach using a two-treatment factorial trial in rural Bangladesh as an example. In this study children <6 months of age were randomly assigned to one of four treatment groups: control, sanitation mobilization, lipid-based nutrient supplementation, and sanitation plus lip-based nutrient supplementation. The design of this study made sample size and study power calculations difficult for two reasons: (1) treatments were deployed at two levels (sanitation mobilization at the community level and lipid supplementation at the individual level); and (2) there were two sources of correlation in the outcome: at the community level and the individual child level.

Generation of the population data set involves: (1) defining the mean and standard deviation of the outcome of interest (for continuous outcomes) or the expected population prevalence (for binary outcomes); (2) defining the number of level 1, 2, 3, ... n units in the population; (3) defining the level 1, 2, 3, ... n variance terms; (4) simulating a population of individuals eligible for sampling based on the specified number of level 1, 2, 3, ... n units; (5) assignment of a value for the outcome variable to each member of the simulated population, and; (6) adjustment of the value of the outcome variable for each individual to account for clustering using the level 1, 2, 3, ... n variance terms.

Code written in the R programming language (6) to generate a population data set for the Reunion Island dairy cow reproduction example (25) and estimate a sample size to meet the requirements of the study is provided in the Supplementary material accompanying this paper.

Figure 2 is an image plot showing the proportion of simulations where the sample estimate of the logarithm of calving to conception interval was within 5% of the true population value as a function of the number of sampled herds and the number of cows sampled from within each herd. In **Figure 2** the superimposed contour line shows the herd-cow sample size combinations where >95% of simulations returned an estimate of the logarithm of calving to conception interval that was within 5% of the true population value, in agreement with the requirement for 2 lactations from 13 cows from 2 herds (n = 52 lactations) calculated using the formula-based approach presented in **Box 8**. When the results of simulations are presented in this way one can appreciate that there is some flexibility in the combinations of herd and cow numbers that need to be sampled to meet the requirements of the study. For example, **Figure 2** shows that the estimate of mean logarithm of calving to conception interval would be within 5% of the true population value if a smaller number of cows (e.g., $n = 6$) were sampled from a larger number of herds (e.g., $n = 6$).

In summary, the process of simulation replaces the time and effort to derive a formula-based approach for a complex study design with basic programming and computer simulation time. An additional positive side effect is that the process of simulation requires investigators to define the structure of their study population, the expected value and variability of the outcome of interest and how the results of the study will be analyzed once the data are collected. This reduces the likelihood of investigators exploring alternative analytical approaches in the presence of negative findings, consistent with CONSORT guidelines (42).

CONCLUSIONS

This paper has provided an overview of the reasons researchers might need to calculate an appropriate sample size in veterinary epidemiology and a summary of different sample size calculation methods. In contrast to human epidemiology individual study subjects in veterinary epidemiology are almost always aggregated into hierarchical groups (43) and, for this reason, sample size estimates calculated using simple formulae that assume independence are usually not appropriate in a veterinary setting. This paper provides details of two approaches for dealing with this problem: (1) inflation of a crude sample size estimate using a design effect; and (2) use of a simulation-based approaches. The key advantage of simulation-based approaches is that appropriate sample sizes can be estimated for complex study designs for which formula-based methods are not available.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets.2020.539573/full#supplementary-material

REFERENCES

1. Detsky A, Sackett D. When was a "negative" clinical trial big enough? How many patients you needed depends on what you found. *Arch Intern Med*. (1985) 145:709–12. doi: 10.1001/archinte.1985.00360040141030
2. Fitzner K, Heckinger E. Sample size calculation and power analysis: a quick review. *Diab Educ*. (2010) 36:701–7. doi: 10.1177/0145721710380791
3. Cook J, Julious S, Sones W, Hampson L, Hewitt C, Berlin J, et al. DELTA(2) guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. *Br Med J*. (2018) 363:k3750. doi: 10.1136/bmj.k3750
4. Delucchi K. Sample size estimation in research with dependent measures and dichotomous outcomes. *Am J Public Health*. (2004) 94:372–7. doi: 10.2105/AJPH.94.3.372
5. Stevenson M, Sergeant E, Nunes T, Heuer C. epiR: tools for the analysis of epidemiological data. In: *Faculty of Veterinary and Agricultural Sciences*. Melbourne, Vic: The University of Melbourne (2020). Available online at: https://cran.rproject.org/web/packages/epiR/index.html
6. R Core Team. R: A language and environment for statistical computing. In: R Foundation for Statistical Computing, Vienna, Austria (2020).
7. Levy P, Lemeshow S. Sampling of Populations Methods and Applications. London: Wiley Series in Probability and Statistics (2008).
8. Kirkwood B, Sterne J. *Essential Medical Statistics*. London: Wiley-Blackwell (2003).
9. Woodward M. *Epidemiology Study Design and Data Analysis*. New York, NY: Chapman & Hall CRC (2014).
10. Dupont W. Power calculations for matched case-control studies. *Biometrics*. (1988) 44:1157–68. doi: 10.2307/2531743
11. Lemeshow S, Hosmer D, Klar J, Lwanga S. *Adequacy of Sample Size in Health Studies*. New York, NY: John Wiley and Sons (1990).
12. Breslow N, Day N. *Statistical Methods in Cancer Research Volume 1: The Analysis of Case-Control Studies IARC Scientific Publication No. 32*. Lyon: International Agency for Research on Cancer (1980).
13. Rose S, Laan M. Why match? Investigating matched case-control study designs with causal effect estimation. *Int J Biostat*. (2009) 5:1. doi: 10.2202/1557-4679.1127
14. Lesaffre E. Superiority, equivalence, non-inferiority trials. *Bull NYU Hosp Joint Dis*. (2008) 66:150–4.
15. Chow S-C, Shao J, Wang H, Lokhnygina Y. *Sample Size Calculations in Clinical Research*. London: CRC Press (2018).
16. Julious S. *Sample Sizes for Clinical Trials*. London: CRC Press.
17. Machin D, Campbell M, Tan S, Tan S. *Sample Size Tables for Clinical Studies*. London: John Wiley & Sons (2009).
18. Head S, Kaul S, Bogers A, Kappetein A. Non-inferiority study design: lessons to be learned from cardiovascular trials. *Eur Heart J*. (2012) 33:1318–24. doi: 10.1093/eurheartj/ehs099
19. Walker E, Nowacki A. Understanding equivalence and noninferiority testing. *J Gen Int Med*. (2011) 26:192–6. doi: 10.1007/s11606-010-1513-8
20. Otte J, Gumm I. Intra-cluster correlation coefficients of 20 infections calculated from the results of cluster-sample surveys. *Prev Vet Med*. (1997) 31:147–50. doi: 10.1016/S0167-5877(96)01108-7
21. Dohoo I, Tillard E, Stryhn H, Faye B. The use of multilevel models to evaluate sources of variation in reproductive performance in dairy cattle in

Reunion Island. *Prev Vet Med*. (2001) 50:127–44. doi: 10.1016/S0167-5877(01)00191-X
22. Martin P, Cameron A, Barfod K, Sergeant E, Greiner M. Demonstrating freedom from disease using multiple complex data sources 2: Case study—Classical swine fever in Denmark. *Prev Vet Med*. (2007) 79:98–115. doi: 10.1016/j.prevetmed.2006.09.007
23. Martin P, Cameron A, Greiner M. Demonstrating freedom from disease using multiple complex data sources 1: a new methodology based on scenario trees. *Prev Vet Med*. (2007) 79:71–97. doi: 10.1016/j.prevetmed.2006.09.008
24. MacDiarmid S. Future options for brucellosis surveillance in New Zealand beef herds. *N Z Vet J*. (1988) 36:39–42. doi: 10.1080/00480169.1988.35472
25. Martin S, Shoukri M, Thorburn M. Evaluating the health status of herds based on tests applied to individuals. *Prev Vet Med*. (1992) 14:33–43. doi: 10.1016/0167-5877(92)90082-Q
26. Cameron A, Baldock F. A new probability formula for surveys to substantiate freedom from disease. *Prev Vet Med*. (1998) 34:1–17. doi: 10.1016/S0167-5877(97)00081-0
27. Cameron A. *Survey Toolbox — A Practical Manual and Software Package for Active Surveillance of Livestock Diseases in Developing Countries*. Canberra: Australian Centre for International Agricultural Research (1999).
28. Cameron A, Baldock F. Two-stage sampling in surveys to substantiate freedom from disease. *Prev Vet Med*. (1998) 34:19–30. doi: 10.1016/S0167-5877(97)00073-1
29. Christensen J, Gardner I. Herd-level interpretation of test results for epidemiologic studies of animal diseases. *Prev Vet Med*. (2000) 45:83–106. doi: 10.1016/S0167-5877(00)00118-5
30. Cornfield J. Randomization by group: a formal analysis. *Am J Epidemiol*. (1978) 108:100–2. doi: 10.1093/oxfordjournals.aje.a112592
31. Donner S, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. *Am J Epidemiol*. (1981) 114:906–14. doi: 10.1093/oxfordjournals.aje.a113261
32. Pagel C, Prost A, Lewycka S, Das S, Colbourn T, Mahapatra R, et al. Intracluster correlation coefficients and coefficients of variation for perinatal outcomes from five cluster-randomised controlled trials in low and middle-income countries: results and methodological implications. *Trials*. (2011) 12:51. doi: 10.1186/1745-6215-12-151
33. Singh J, Liddy C, Hogg W, Taljaard M. Intracluster correlation coefficients for sample size calculations related to cardiovascular disease prevention and management in primary care practices. *BMC Res Notes*. (2015) 8:89. doi: 10.1186/s13104-015-1042-y
34. Donner A. A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. *Int Stat Rev*. (1986) 54:67–82. doi: 10.2307/1403259
35. Ridout M, Demétrio C, Firth D. Estimating intraclass correlation for binary data. *Biometrics*. (1999) 55:137–48. doi: 10.1111/j.0006-341X.1999.00137.x
36. Searle S. A biometrics invited paper. Topics in variance component estimation. *Biometrics*. (1971) 27:1–76. doi: 10.2307/2528928
37. Goldstein H, Browne W, Rasbash J. Partitioning variation in multilevel models. *Underst Stat*. (2002) 1:223–32. doi: 10.1207/S15328031US0104_02
38. Eldridge S, Kerry S. *A Practical Guide to Cluster Randomised Trials in Health Services Research*. London: John Wiley & Sons (2012).
39. Eldridge S, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *Int J Epidemiol*. (2006) 35:1292–300. doi: 10.1093/ije/dyl129
40. Rutterford C, Copas A, Eldridge S. Methods for sample size determination in cluster randomized trials. *Int J Epidemiol*. (2015) 44:1051–67. doi: 10.1093/ije/dyv113
41. Arnold B, Hogan D, Colford J, Hubbard A. Simulation methods to estimate design power: an overview for applied research. *BMC Med Res Methodol*. (2011) 11:94. doi: 10.1186/1471-2288-11-94
42. Schulz K, Altman D, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Br Med J*. (2010) 340:c332. doi: 10.1136/bmj.c332
43. Stevenson M, Firestone S. Veterinary epidemiology and 'One Health' in Australasia. *Australas Epidemiol*. (2015) 22:35–38.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership