frontiers | Research Topics

Lessons and policy consequences of mathematical modelling in relation to ongoing pandemics

Edited by
Theodore Gyle Lewis, Pierpaolo Ferrante and Waleed Isa Al Mannai

Published in
Frontiers in Public Health

ISSN 1664-8714
ISBN 978-2-8325-3770-1
DOI 10.3389/978-2-8325-3770-1

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

Lessons and policy consequences of mathematical modelling in relation to ongoing pandemics

Topic editors

Theodore Gyle Lewis — Naval Postgraduate School, United States

Pierpaolo Ferrante — National Institute for Insurance against Accidents at Work (INAIL), Italy

Waleed Isa Al Mannai — New York Institute of Technology Bahrain, Bahrain

Citation

Lewis, T. G., Ferrante, P., Mannai, W. I. A., eds. (2023). *Lessons and policy consequences of mathematical modelling in relation to ongoing pandemics*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3770-1

Table of contents

05 Editorial: Lessons and policy consequences of mathematical modeling in relation to ongoing pandemics
Pierpaolo Ferrante

08 The first 2 years of COVID-19 in Italy: Incidence, lethality, and health policies
Pierpaolo Ferrante

22 Impact of early phase COVID-19 precautionary behaviors on seasonal influenza in Hong Kong: A time-series modeling approach
Chun-Pang Lin, Ilaria Dorigatti, Kwok-Leung Tsui, Min Xie, Man-Ho Ling and Hsiang-Yu Yuan

32 The role of models as a decision-making support tool rather than a guiding light in managing the COVID-19 pandemic
Adi Niv-Yagoda, Royi Barnea and Efrat Rubinshtein Zilberman

41 Measuring the impact of the COVID-19 epidemic on university resumption and suggestions for countermeasures
Shi Yin, Lijun Ma, Tong Dong and Ying Wang

50 Two-year follow-up of the COVID-19 pandemic in Mexico
Antonio Loza, Rosa María Wong-Chew, María-Eugenia Jiménez-Corona, Selene Zárate, Susana López, Ricardo Ciria, Diego Palomares, Rodrigo García-López, Pavel Iša, Blanca Taboada, Mauricio Rosales, Celia Boukadida, Alfredo Herrera-Estrella, Nelly Selem Mojica, Xaira Rivera-Gutierrez, José Esteba Muñoz-Medina, Angel Gustavo Salas-Lais, Alejandro Sanchez-Flores, Joel Armando Vazquez-Perez, Carlos F. Arias and Rosa María Gutiérrez-Ríos

64 Intensity and lag-time of non-pharmaceutical interventions on COVID-19 dynamics in German hospitals
Yvette Montcho, Paul Klingler, Bruno Enagnon Lokonon, Chénangnon Frédéric Tovissodé, Romain Glèlè Kakaï and Martin Wolkewitz

72 Content and sentiment surveillance (CSI): A critical component for modeling modern epidemics
Shi Chen, Shuhua Jessica Yin, Yuqi Guo, Yaorong Ge, Daniel Janies, Michael Dulin, Cheryl Brown, Patrick Robinson and Dongsong Zhang

78 Digital cities and the spread of COVID-19: Characterizing the impact of non-pharmaceutical interventions in five cities in Spain
Jorge P. Rodríguez, Alberto Aleta and Yamir Moreno

89 A general modeling framework for quantitative tracking, accurate prediction of ICU, and assessing vaccination for COVID-19 in Chile
Patricio Cumsille, Oscar Rojas-Díaz and Carlos Conca

106 Lag-time effects of vaccination on SARS-CoV-2 dynamics in German hospitals and intensive-care units
Bruno Enagnon Lokonon, Yvette Montcho, Paul Klingler, Chénangnon Frédéric Tovissodé, Romain Glèlè Kakaï and Martin Wolkewitz

115 The role of machine learning in health policies during the COVID-19 pandemic and in long COVID management
Lindybeth Sarmiento Varón, Jorge González-Puelma, David Medina-Ortiz, Jacqueline Aldridge, Diego Alvarez-Saravia, Roberto Uribe-Paredes and Marcelo A. Navarrete

127 What we talk about when we talk about COVID-19 vaccination campaign impact: a narrative review
Horácio N. Hastenreiter Filho, Igor T. Peres, Lucas G. Maddalena, Fernanda A. Baião, Otavio T. Ranzani, Silvio Hamacher, Paula M. Maçaira and Fernando A. Bozza

138 Keeping kids in school: modelling school-based testing and quarantine strategies during the COVID-19 pandemic in Australia
Romesh G. Abeysuriya, Rachel Sacks-Davis, Katherine Heath, Dominic Delport, Fiona M. Russell, Margie Danchin, Margaret Hellard, Jodie McVernon and Nick Scott

149 Socioeconomic determinants of stay-at-home policies during the first COVID-19 wave
Pablo Valgañón, Unai Lería, David Soriano-Paños and Jesús Gómez-Gardeñes

159 A study on the emotional and attitudinal behaviors of social media users under the sudden reopening policy of the Chinese government
Qiaohe Zhang, Tianyue Niu, Jinhua Yang, Xiaochen Geng and Yinhuan Lin

frontiers | Frontiers in Public Health

TYPE Editorial
PUBLISHED 09 October 2023
DOI 10.3389/fpubh.2023.1281493

OPEN ACCESS

EDITED AND REVIEWED BY
Marc Jean Struelens,
Université Libre de Bruxelles, Belgium

*CORRESPONDENCE
Pierpaolo Ferrante
✉ p.ferrante@inail.it

RECEIVED 22 August 2023
ACCEPTED 22 September 2023
PUBLISHED 09 October 2023

CITATION
Ferrante P (2023) Editorial: Lessons and policy consequences of mathematical modeling in relation to ongoing pandemics.
Front. Public Health 11:1281493.
doi: 10.3389/fpubh.2023.1281493

Editorial: Lessons and policy consequences of mathematical modeling in relation to ongoing pandemics

Pierpaolo Ferrante*

Department of Occupational and Environmental Medicine, Epidemiology and Hygiene, Italian National Workers' Compensation Authority (INAIL), Rome, Italy

KEYWORDS

COVID-19 mathematical modeling, COVID-19 policy, global pandemic plan, pandemic containment measures, pandemic preparedness

Editorial on the Research Topic
Lessons and policy consequences of mathematical modeling in relation to ongoing pandemics

COVID-19 marked the second pandemic of the 21st century, following the swine pandemic (A/H1N1pdm09) of 2009. It is the third outbreak of coronaviruses after SARS (2003) and MERS (2013), and the seventh monitored outbreak (including those caused by the Zika virus and the avian flus A/H5N1 and A/H7N9). Excluding the Zika virus (belonging to the "Flaviviridae" family) and A/H5N1, the remaining were emerging viruses (1). While some uncertainty persists about the origin of SAR-CoV-2, current evidence suggests an animal origin (2). According to numerous studies highlighting the role of Climate change in increasing the risk rate of cross-species viral transmission (3), the potential for pandemics to emerge as one of the most significant threats to humanity in the future is evident (4). The development of coordinated national pandemic plans should be a priority for every country in order to release a global response to a global issue (5).

An effective pandemic plan should be designed through a multidisciplinary approach, offering flexibility for calibration based on evolving data evidence, and structured around the following points:

(a) Epidemiological: Establish a robust epidemiological surveillance system encompassing the entire national territory. Based on varying assumptions about the virus's virulence, simulated scenarios should be run to project the virus's potential spread within the population. Epidemiological thresholds to activate restrictive public policies (including mobility restrictions and social distancing) as well as mandatory sectorial behaviors (including the use of FFP2 face mask in public transport, smart working, and distance learning) should be proposed based on predictive model outputs. During the pandemic, it is important to adapt the most relevant existing models to data evidence and develop new models as necessary.

(b) Medical: Research and development of vaccines and treatments to address the biological aspects of preventing and caring the disease. Healthcare workers must be trained to manage the pandemic with simulations over time to ensure basic preparedness.

(c) Logistics: Implementing the healthcare response by defining the necessary resources (including diagnostic tests, protective equipment, hospital capacity, and vaccines) and organizing their distribution across territories. Coordination from national government to local institutions with multiple decision-making centers should be implemented to facilitate collective territorial coherence and minimize the possible consequences of government crisis. Primary health care should play an active role in guaranteeing system resilience (6).

(d) Political and ethical: Selecting the epidemiological thresholds and determining their implementation pattern requires a delicate balance among competing human rights, including liberty, economy, and health. Sociologists, jurists, and constitutionalists should engage in relevant discussions aimed at increasing the likelihood of public acceptance.

(e) Communication: Design a communication campaign using multimedia platforms to effectively convey accessible and clear information in a visually and verbally engaging manner, reaching a wide audience. Emphasize scientific dissemination rather than TV entertainment programs and address the anti-vax problem with the appropriate information.

This Research Topic includes 11 original research, two brief reports, two reviews, and one perspective paper. All the original research and brief reports either used publicly available data or were made accessible upon request to the authors. Four studies included the analysis code. Simulations were carried out through the COVASIM model (7), which is implemented in free python code.[1] Focusing on mathematical modeling, the collected papers addressed points (a) and (e) of the previous list during the COVID-19 pandemic.

The primary overarching conclusion from this Research Topic is the remarkable proliferation of mathematical modeling during the emergency period. This global effort reflects the impressive mobilization of human societies worldwide as well as an underlying lack of preparedness and coordination. For instance, the Israeli health response relied on three different models with varying assumptions and outcomes (Niv-Yagoda et al.).

An overview of pandemic characteristics was provided through descriptive and predictive models. By analyzing cases from the Mexico's surveillance system during the first 2 years, Loza et al. confirmed the key role of comorbidities in disease severity and the effectiveness of vaccination campaigns. Ferrante introduced the negative binomial model to estimate the incidence of infection from mortality in Italy. Results indicate that over 40% of infections went undetected, with the majority occurring before the introduction of rapid tests. Cumsille et al. predicted the occupancy of intensive care units by adding to the SIR model a compartment representing the number of patients in intensive care and two parameters describing the rate from susceptible to recovered (due to vaccine protection) and the vice versa (due to vaccine immunity decay).

A description of the models used to evaluate the effectiveness of the vaccination campaign along with two study using them are included. Filho et al. conducted a review on studies addressing the impact of a vaccination program. They found that half of them simulated scenarios with and without vaccines, while the others compared the populations before and after vaccination. By simulating the scenario without a vaccination, Ferrante found that vaccines prevented 115,000 deaths during the first two pandemic years in Italy. By comparing the pre- and post-vaccination populations, Lokonon et al. studied the lag-time effects of vaccination through a quasi-Poisson regression with a distributed lag linear model. They found a significant protective effect when the 40% of people were vaccinated, with a lag time of 15 days for the effect of the third dose.

Non-pharmaceutical interventions were extensively investigated, including their impact on seasonal influenza. Montcho et al. analyzed these interventions using a distributed lag linear model. They found that stricter restrictions led to fewer admissions in regular and intensive care units, with a 9–10 day time lag. Rodríguez et al. investigated the impact of non-pharmaceutical interventions in Spain using a data-driven agent-based model. Simulations revealed that the combination of tracing and testing, along with the associated isolation of positive individuals, halved infections and deaths. Valgañón et al. investigated the socioeconomic determinants of stay-at-home through a SEIR model that included a permeability parameter and a predeceased compartment. Their study highlighted the need for equitable global policies, showing the challenges low-income countries face in mitigating the virus spread and protecting vulnerable populations. Lin et al. studied the effects of non-pharmaceutical policies on the seasonal flu and found that wearing face masks and avoiding crowded places protected ∼20 and 40% of people, respectively. Furthermore, if more than 85% of people had adopted both behaviors the reproduction number could have been <1.

As with other respiratory viruses, schools played a relevant role in the COVID-19 spread. Yin et al. studied the university resumption impact using a disaster management perspective and the pressure–state–response model. Their model included six factors representing disaster hazards that university can only monitor (including epidemic risk level of the school's location and means of transportation back to school) and fourteen factors related to system vulnerability that can also be controlled (including student behaviors and routine campus activities). Through simulations, Abeysuriya et al. compared three testing strategies in schools: home quarantine of all contacts of a positive case; "test-to-stay" strategy for close contacts of a case for 7 days; and an asymptomatic surveillance strategy involving twice-weekly screening of all students. Compared to extended home quarantine, test-to-stay strongly increased days of face-to-face teaching while maintaining a similar effectiveness for reducing school infections. Asymptomatic screening was beneficial in reducing both infections and lost days of face-to-face teaching especially when community prevalence was high.

The COVID-19 pandemic also marked an intensive use of machine learning and sentiment analysis in epidemiological modeling. Varón et al. performed a review to describe the role of machine learning in health policies. They found an increasing usage of these methods both in COVID-19 and long COVID studies, including clinical diagnosis, epidemiological analysis, drug discovery, patterns and relationships of symptoms, and predicting risk indicators. Chen et al. proposed integrating epidemic modeling with content and sentiment infoveillance based on natural language

1 Available online at: https://github.com/InstituteforDiseaseModeling/covasim.

processing. They concluded that infoveillance from massive social media data complements and enhances current epidemic models. Zhang et al. conducted a sentiment analysis of the Chinese reopening policy after 3 years of "zero-COVID" measures. They found a negative attitude toward "sudden" measures and suggest preparing people in advance with relevant health consultation services and an effective communication strategy.

Results from this Research Topic suggest several points to upgrade the pandemic response from national to a global level. In particular, for improving the initial preparedness we recommend:

1) WHO should lead research efforts to identify and classify potential spillover viruses (8) and advance research on related vaccines and therapies.
2) WHO should develop guidelines for establishing public and standardized national-level epidemiological virus surveillance systems based on Statistical Data and Metadata eXchange (9). These systems ensure the collection of consistent national data, which can be simply transmitted to a global database (such as the global influenza surveillance and Response System) and made accessible through user-friendly APIs.
3) National pandemic plans should include simulations of virus spread, considering varying levels of virulence and transmission abilities. These simulations should use an agent-based model and encompass factors such as the saturation level of hospitals and the impact of pharmaceutical and non-pharmaceutical policies.
4) Countries should integrate pandemic-era hygiene rules (such as wearing face mask in crowded places and washing hands after touching surfaces potentially contaminated) into primary school hygiene education.

Additionally, priorities during a pandemic should include:

5) In countries initially affected, lethality should be promptly estimated through community serosurveys focused on the area surrounding the initial deaths. Subsequent refinement can be achieved through regional and national serosurveys that track virus circulation.
6) Estimating the lethality hazard ratios for the virus variants compared to the original strain, along with the relative risks of infection and death among vaccinated and unvaccinated individuals, will allow for the consistent application of the negative binomial model throughout the entire pandemic period.
7) Ensure the rapid availability of reliable rapid tests to support a worldwide testing campaign for identifying positive cases and contact tracing.
8) Secure the swift availability and equitable global distribution of effective vaccines.
9) Calibrate health policy and communication campaigns based on sentiment analysis to increase public compliance.

Author contributions

PF: Conceptualization, Writing—original draft, Writing—review and editing.

Funding

This research was supported and founded by INAIL (the Italian Workers Compensation Authority).

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.

Publisher's note

References

1. Piret J, Boivin G. Pandemics throughout history. *Front Microbiol.* (2021) 11:631736. doi: 10.3389/fmicb.2020.631736

2. Gostin LO, Gronvall GK. The origins of Covid-19—why it matters (and why it doesn't). *N Engl J Med.* (2023) 388:2305–8. doi: 10.1056/NEJMp2305081

3. Carlson CJ, Albery GF, Merow C, Trisos CH, Zipfel CM, Eskew EA, et al. Climate change increases cross-species viral transmission risk. *Nature.* (2022) 607:555–62. doi: 10.1038/s41586-022-04788-w

4. Majedul Islam MM. Threats to humanity from climate change. In: Bandh SA, editor. *Climate Change: The Social and Scientific Construct*. Cham: Springer International Publishing (2022). p. 21—36. doi: 10.1007/978-3-030-86290-9_2

5. Medicine. Future pandemics: failing to prepare means preparing to fail. *Lancet.* (2022) 10.3: 221. doi: 10.1016/S2213-2600(22)00056-X

6. Mathews M, Ryan D, Hedden L, Lukewich J, Marshall EG, Buote R, et al. Strengthening the integration of primary care in pandemic response plans: a qualitative interview study of Canadian family physicians. *Br J Gen Pract.* (2023) 73:e348–55. doi: 10.3399/BJGP.2022.0350

7. Kerr CC, Stuart RM, Mistry D, Abeysuriya RG, Rosenfeld K, Hart GR, et al. Covasim: an agent-based model of COVID-19 dynamics and interventions. *PLoS Comput Biol.* (2021) 17:e1009149. doi: 10.1371/journal.pcbi.1009149

8. Mollentze N, Streicker DG. Predicting zoonotic potential of viruses: where are we? *Curr Opin Virol.* (2023) 61:101346. doi: 10.1016/j.coviro.2023.101346

9. Bender S, Blaschke J, Hirsch C. "Statistical data production in a digitized age: the need to establish successful workflows for micro data access," In: Snijkers G, Bavdaž M, Bender S, Jones J, MacFeeley S, Sakshaug JW, Thompson KJ, van Delden A, editors. *Advances in Business Statistics, Methods and Data Collection, Chapter 22*, Hoboken: Wiley. (2023) doi: 10.1002/9781119672333.ch22

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 01 November 2022
DOI 10.3389/fpubh.2022.986743

OPEN ACCESS

EDITED BY
Maria Rosario O. Martins,
New University of Lisbon, Portugal

REVIEWED BY
Giuseppe De Natale,
National Institute of Geophysics and Volcanology (INGV), Italy
Arianna Calistri,
University of Padua, Italy

*CORRESPONDENCE
Pierpaolo Ferrante
p.ferrante@inail.it

SPECIALTY SECTION
This article was submitted to Infectious Diseases - Surveillance, Prevention and Treatment, a section of the journal Frontiers in Public Health

RECEIVED 06 July 2022
ACCEPTED 21 September 2022
PUBLISHED 01 November 2022

CITATION
Ferrante P (2022) The first 2 years of COVID-19 in Italy: Incidence, lethality, and health policies.
Front. Public Health 10:986743.
doi: 10.3389/fpubh.2022.986743

The first 2 years of COVID-19 in Italy: Incidence, lethality, and health policies

Pierpaolo Ferrante *

Department of Occupational and Environmental Medicine, Epidemiology and Hygiene, Italian National Workers' Compensation Authority (INAIL), Rome, Italy

Background: The novel coronavirus disease 2019 (COVID-19) is an ongoing pandemic that was first recognized in China in December 2019. This paper aims to provide a detailed overview of the first 2 years of the pandemic in Italy.

Design and methods: Using the negative binomial distribution, the daily incidence of infections was estimated through the virus's lethality and the moving-averaged deaths. The lethality of the original strain (estimated through national sero-surveys) was adjusted daily for age of infections, hazard ratios of virus variants, and the cumulative distribution of vaccinated individuals.

Results: From February 24, 2020, to February 28, 2022, there were 20,833,018 (20,728,924–20,937,375) cases distributed over five waves. The overall lethality rate was 0.73%, but daily it ranged from 2.78% (in the first wave) to 0.15% (in the last wave). The first two waves had the highest number of daily deaths (about 710) and the last wave showed the highest peak of daily infections (220,487). Restriction measures of population mobility strongly slowed the viral spread. During the 2nd year of the pandemic, vaccines prevented 10,000,000 infections and 115,000 deaths.

Conclusion: Almost 40% of COVID-19 infections have gone undetected and they were mostly concentrated in the first year of the pandemic. From the second year, a massive test campaign made it possible to detect more asymptomatic cases, especially among the youngest. Mobility restriction measures were an effective suppression strategy while distance learning and smart working were effective mitigation strategies. Despite the variants of concern, vaccines strongly reduced the pandemic impact on the healthcare system avoiding strong restriction measures.

KEYWORDS

COVID-19, incidence, lethality, health policy, negative binomial, moving averages

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a new virus identified in Wuhan (Hubei, China) in late 2019 (1). SARS-CoV-2 causes the coronavirus disease 2019 (COVID-19), an illness that ranges from mild flu symptoms to bilateral interstitial pneumonia (2). The virus spread so quickly around the world that the World Health Organization (WHO) declared the COVID-19 a Public Health Emergency of International Concern on January 30, 2020 and a pandemic on

March 11, 2020, (3). Unlike other coronaviruses, the SARS-CoV-2 is able to spread through pre- and asymptomatic infections that are difficult to detect and isolate, requiring health authorities to test all contacts of confirmed cases to lower the risk of spread (4, 5). The lethality of the original strain was estimated using infection fatality ratios (IFR) assessed through several national sero-surveys (6, 7). While relatively low in the whole population (<1 death per 100 infections in developed countries), the risk of death is shown to increase with age (up to 10–15 deaths per 100 infections in people aged more than 75 years) and in patients who are immunosuppressed or have concomitant comorbidities (8, 9). Furthermore, since the prognosis of severe cases depends on the availability of intensive care beds, lethality increases when critical care capacity is saturated (10). To address the pandemic, a global vaccination campaign was launched, and pharmaceutical industry developed candidate COVID-19 vaccines at an unprecedented speed. By the end of 2020, global Medicines Agencies had conditionally approved several vaccines based on different technologies, with others close behind (11, 12). During the first 2 years of the pandemic (since December 29, 2019, to February 28, 2022), National Health Institutions detected 444,900,763 confirmed cases and reported 6,020,752 deaths worldwide[1]. The highest number of infections favored mutations in the viral genome sequence and led to generation and spread of many viral variants (13). WHO coordinates national and subnational research aimed at sequencing RNA viral genomes detected in infected people and classifying variants of concern (VOC) that may pose a greater risk to global public health[2]. From May 2020 to February 2022, WHO identified five consecutive VOCs: Alpha, Beta, Gamma, Delta, and Omicron. Each variant showed an increased capacity to spread (even within vaccinated people) and although the debate on virulence is still open, it would appear that all the VOCs except Omicron caused a disease with higher severity and mortality (14, 15). In February 2020, Italy was the European epicenter of the SARS-CoV-2 spreading. The unexpectedly high speed of transmissions quickly resulted in hospital saturation and forced the Italian government to establish a national lockdown. Restriction measures blocked the first wave and were gradually removed in parallel with the development of a robust COVID-19 contact tracing system. To avoid lockdown during the second wave, the national government has applied a standard set of restriction measures (from soft to hard) at the regional level based on the risk of spread evaluated on a weekly basis. The risk level by geographic area (represented by a colored map: white = low, yellow = moderate, orange = high, red = highest) was evaluated by determining weekly estimates of incidence and reproductive number (R_t). During the second wave (December 27th, 2020), a national vaccination campaign was launched using two messenger RNA (Pfizer-BioNTech, Moderna) and two vector vaccines (Janssen, Vaxzevria) (14). Given the high percentage of vaccinated people in the third wave, the hospital saturation levels replaced the R_t in the risk evaluation. Although vaccine protection declined over time (especially against virus variants), protection returned following administration of the booster dose especially against the development of severe infections (16–19). Health institution recommended a booster shot after 4 months from the standard cycle in September 2021, and included children aged 5–11 years in the vaccine campaign in December 2021 (19, 20). This study aims to provide a detailed overview of the first 2 years of the pandemic in Italy, where 13,000,000 of confirmed cases and 155,000 deaths were reported from February 2020 to 2022. The current paper is part of a larger project aimed at describing the epidemiology of Italian COVID-19 pandemic and follows an initial article introducing the method used to describe the pandemic in its 1st year (21).

Methods

Study design

This study analyzed public data of COVID-19 in Italy collected in the national registry by the Civil Protection (CP) and the National Health Institute (ISS).

Settings

The Italian Government declared a health emergency status on January 3, 2020 and extended it to March 31, 2022. The CP was delegated to manage the process and established a system to collect COVID-19 data in a national registry (managed by the ISS). Aggregate data on incidence and vaccination are published daily. The ISS reviews and updates the registry data to account for data reporting delays and regional recounts and releases an updated report with details including the age distribution of detected cases. The European Center for Disease Prevention and Control (ECDC) collects VOCs continental data through the European Surveillance System (TESSy).

Participants

All confirmed cases of COVID-19 in Italy.

Outcomes

The primary outcomes were: (1) the number $\mathbf{N_k}$ of persons who became infected on the kth day of pandemic; (2) the number $\mathbf{D_k}$ of persons who died (over time) among $\mathbf{N_k}$ (i.e., the

1 https://covid19.who.int (accessed September 19, 2022).

2 https://www.who.int/en/activities/tracking-SARS-CoV-2-variants (accessed September 19, 2022).

number of deaths by the infection day); (3) the number $\upsilon_{\mathbf{k}}$ of persons who were officially detected among $\mathbf{N_k}$ (i.e., the number of diagnosed cases by the infection day).

Data sources/measurement

Aggregate data from the national COVID-19 registry and the vaccine campaign are stored in public repositories and updated daily. The data include daily counts of performed tests, of diagnosed cases and fatalities who tested positive using the polymerase chain reaction or the rapid antigen test (beginning on January 8, 2021), and of persons who received vaccine shots by region[3]. The ISS provides a weekly report that includes the median age of detected cases, estimates of vaccine protection and (beginning on December 7, 2020) the distribution of detected cases by 10-year age class[4] (22). The ECDC releases European data on VOCs[5], the National Institute of Statistics releases data from the sero-survey (May 25–July 15, 2020)[6] and on Italian population[7].

Statistical analysis

As already highlighted by De Natale et al. at the onset of the pandemic, the high number of asymptomatic infections makes deaths more suitable than detected cases for estimating incidence (23). Given the probability p_k of dying after having caught the infection on the kth pandemic day ($k \in \mathbb{Z}^+$), we used the negative binomial distribution to estimate the daily number of infections (N_k) from the resulting deaths over time [D_k; Section Modeling the Incidence of Infections (Negative Binomial Distribution)]. First, we estimated D_k by applying the weighted moving average to deaths (that are recorded by the occurrence date, Section Estimating D_k and υ_k: Weighted Moving Average). Second, we modeled the probability p_k accounting for the age at infection, VOCs prevalence and population vaccination level (Section Modeling the Daily Probability to Die p_k). Using other simple assumptions, we evaluated excess death (for health system saturation) and lives saved by vaccines (Sections Excess Death and Vaccine Effect). Finally, we used the number of detected cases υ_k among N_k to check the admissibility ($N_k > \upsilon_k$) of estimates (Sections Estimating D_k and υ_k: Weighted Moving Average and Checking Estimates). In the following, we will proceed with the mathematical formulation, which will be progressively upgraded, in the next sections, to consider the more complex probabilities involving age classes, different strains, and vaccination level. Once the main formulas are established, the estimated variables used to determine the solutions will be given in the Section Estimating the Daily Lethality.

Modeling the incidence of infections (negative binomial distribution)

Let $X_k^{(j)}$ be the binary random variable representing the outcome (1 = dead; 0 = recovered) of the jth person infected on the kth day of the pandemic

$$X_k^{(j)} = \begin{cases} 1 & p_k \\ 0 & 1 - p_k \end{cases}$$

and let $N_k^{(D_k)}$ be the random variable representing the rank of the daily infection resulting in the D_k-th death, the probability of $N_k^{(D_k)}$ follows a negative binomial distribution with parameters D_k and p_k

$$\begin{aligned} P\left\{N_k^{(D_k)} = n\right\} &= P\left\{\sum_{j=1}^{n-1} X_k^{(j)} = D_k - 1,\right. \\ &\qquad \left.\sum_{j=1}^{n} X_k^{(j)} = D_k\right\} \\ &= \binom{n-1}{D_k - 1} p_k^{D_k} \left(1 - p_k\right)^{n - D_k} \end{aligned} \tag{1}$$

with $k, D_k \in Z^+, n \geq D_k$. We estimated the number of daily infections (with the related 95% CI) as the mean of Equation (1)

$$\widehat{N}_k^{(D_k)} = E\left[N_k^{(D_k)}\right] = \frac{D_k}{p_k}. \tag{2}$$

Estimating D_k and υ_k: Weighted moving average

Let $d_{k,k+j}$ and $\upsilon_{k,k+j}$ be the number of persons infected on the kth pandemic day who died or were diagnosed j days after the infection, the number of deaths (D_k), and detected cases (V_k) among infections on the kth pandemic day can be evaluated as

$$D_k = \sum\nolimits_j d_{k,k+j} \text{ and } V_k = \sum_j \upsilon_{k,k+\,j}.$$

Since only the corresponding number of events by the occurrence date (of death or diagnosis) is available

$$d_{\cdot,k+j} = \sum\nolimits_i d_{i,k+j} \text{ and } \upsilon_{\cdot,k+j} = \sum\nolimits_i \upsilon_{i,k+j} \tag{3}$$

D_k and V_k were estimated as

$$D_k = \sum\nolimits_j \pi_j^{(k+j)} d_{\cdot,k+j} \text{ and } \quad V_k = \sum\nolimits_j \theta_j^{(k+j)} \upsilon_{\cdot,k+j}, \tag{4}$$

3 https://github.com/italia/covid19-opendata-vaccini/tree/master/dati (accessed September 09, 2022).

4 https://github.com/floatingpurr/covid-19_sorveglianza_integrata_italia/tree/main/data (accessed September 09, 2022).

5 https://www.ecdc.europa.eu/en/publications-data/data-virus-variants-covid-19-eueea (accessed September 09, 2022).

6 https://www.istat.it/it/archivio/242676 (accessed September 09, 2022).

7 https://demo.istat.it (accessed September 09, 2022).

where $\pi_j^{(k+j)}$ and $\theta_j^{(k+j)}$ are the fractions

$$\pi_j^{(k+j)} = \frac{d_{k,k+j}}{d_{.,k+j}} \text{ and } \theta_j^{(k+j)} = \frac{\upsilon_{k,k+j}}{\upsilon_{.,k+j}}.$$

Let T_{dead} and T_{diagn} represent the time from infection to death and diagnosis, respectively, and α_k and β_k be the binary variables representing the events to die ($\alpha_k = 1$) or be alive ($\alpha_k = 0$) and to be diagnosed ($\beta_k = 1$) or undetected ($\beta_k = 0$) on the kth pandemic day, $\pi_j^{(k+j)}$ and $\theta_j^{(k+j)}$ can be expressed as the conditional probability to die or be diagnosed j days after the infection

$$\begin{aligned} \pi_j^{(k+j)} &= P\left\{j \le T_{dead} < j+1 | \alpha_{k+j} = 1\right\} \text{ and} \\ \theta_j^{(k+j)} &= P\left\{j \le T_{diagn} < j+1 | \beta_{k+j} = 1\right\}. \end{aligned} \quad (5)$$

The ISS provided estimated quartiles ($Q1$, $Q2$, and $Q3$) of the time distributions from symptoms to death and diagnosis during three different periods (March-May/2020, June-September/2020, and October/2020-December/2020). The ISS estimates for time to death are admissible under symmetric distributions except during the summer period [where there was strong bias from clusters of vacationers (24)]. These biased estimates were not considered and the remaining, which are equivalent [Table 1 in (21)], were extended to the whole studied period. We added 5 days [the mean time from infection to symptoms (25)] to ISS estimates to obtain the corresponding parameters of the probability density function of the time from infection to death and diagnosis

$$\begin{aligned} f_{T_{dead}}^{(\alpha_k=1)}(t) &= \frac{d}{dt} P\left\{T_{dead} < t | \alpha_k = 1\right\} \text{ and} \\ f_{T_{diagn}}^{(\beta_k=1)}(t) &= \frac{d}{dt} P\left\{T_{diagn} < t | \beta_k = 1\right\}. \end{aligned} \quad (6)$$

If necessary, we adjusted for symmetry by replacing the median with the center of first and third quartile and assumed that the functions in Equation (6) follow the truncated normal distribution

$$F_T(t) = \frac{\frac{e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}}{\int_0^{2\mu} \frac{e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}\, dt} \text{ with } t \in [0, 2\mu], \quad (7)$$

where μ and σ are the mean and standard deviation of the parent general normal probability with $\mu = \frac{Q_3+Q_1}{2}$ and $\sigma = \frac{Q_3-Q_1}{1.34896}$. Of note, the Equation (4) with probabilities Equation (5) derived from Equation (7) can be also interpreted as a weighted moving average of period $2\mu + 1$ on time series $d_{.,k+j}$ and $\upsilon_{.,k+j}$ in (3)

$$D_k = \sum_{j=0}^{2\mu} \pi_j^{(k+j)} d_{.,k+j} \text{ and } V_k = \sum_{j=0}^{2\mu} \theta_j^{(k+j)} \upsilon_{.,k+j}.$$

Modeling the daily probability to die p_k

Let $X_{j,\xi,V}$ and $Y_{k,j,\xi,V}$ be the binary random variables representing the events "to die after the infection" and "to be infected on the kth pandemic day," respectively, by 10-year age class (j: 0–9, 10–19, …, 80–89, 90+ years), VOC (ξ: 0 = original strain; 1 = Alpha; 2 = Beta; 3 = Gamma;4 = Delta; 5 = Omicron), and vaccination level (V: 0 = unvaccinated; 1 = uncompleted basic cycle; 2 = completed basic cycle more than 4 months ago; 3 = completed basic cycle in the last 4 months; 4 = received a booster shot). By assuming that the conditional probability $p_{k,j,\xi,V}$ to die after having caught the infection on the kth day does not depend on k, we have that

$$p_{j,\xi,V} = P\left\{X_{j,\xi,V} = 1 | Y_{k,j,\xi,V} = 1\right\} \ \forall\, k \in \mathbb{Z}^+. \quad (8)$$

Let $N_{k,j,\xi,V}$ be the number of infected people on the kth pandemic day by age class, VOC, and vaccination level and $N_{k,.,.,.}$ be the total number of infections on the same day, the overall probability p_k to die among infections on the k-th day is equal to

$$p_k = \sum_j \sum_\xi \sum_V p_{j,\xi,V} \frac{N_{k,j,\xi,V}}{N_{k,.,.,.}}. \quad (9)$$

Now, let $RR_{j,\xi,V}$ be the risk ratio to die of people with the vaccination level V (= 0, 1, 2, 3, 4) vs. unvaccinated ($V = 0$) by age class and VOC

$$RR_{j,\xi,V} = \frac{p_{j,\xi,V}}{p_{j,\xi,0}}, \quad (10)$$

and let $N_{k,j,\xi,.}$ and $N_{k,j,.,.}$ be the number of infections on the kth pandemic, respectively, by age class and VOC (with any vaccination level) and by age class (with any VOC and vaccination level), the Equation (9) can be rewritten through the Equation (10) as

$$p_k = \sum_j \left[\sum_\xi \left(\sum_V RR_{j,\xi,V} \frac{N_{k,j,\xi,V}}{N_{k,j,\xi,.}} \right) p_{j,\xi,0} \frac{N_{k,j,\xi,.}}{N_{k,j,.,.}} \right] \frac{N_{k,j,.,.}}{N_{k,.,.,.}}. \quad (11)$$

Finally, let $S_{j,\xi,0}(t)$ and $S_{j,0,0}(t)$ be the distributions of survival time of unvaccinated people in the jth age class, respectively, for the VOC ξ (= 0, 1, 2, 3, 4, 5) and original virus strain ($\xi = 0$), under the assumption of proportional hazards we have that

$$\frac{d}{dt} log\left[S_{j,\xi,0}(t)\right] = hR_{j,\xi,0} \frac{d}{dt} log\left[S_{j,0,0}(t)\right], \quad (12)$$

where $hR_{j,\xi,0}$ are the hazard ratios by VOC (ξ = 0, 1, 2, 3, 4, 5 vs. $\xi = 0$) by age class for unvaccinated people ($V = 0$). By integrating the Equation (12) over the whole pandemic period, we obtain the following identity

$$S_{j,\xi,0} = \left[S_{j,0,0}\right]^{hR_{j,\xi,0}} \quad (13)$$

and since $S_{j,\xi,0} = 1 - p_{j,\xi,0}$ and $S_{j,0,0} = 1 - p_{j,0,0}$ the Equation (11) can be expressed as

$$p_k = \sum_j \left\{ \sum_\xi \left(\sum_V RR_{j,\xi,V} \frac{N_{k,j,\xi,V}}{N_{k,j,\xi,.}} \right) \left[1 - \left(1 - p_{j,0,0} \right)^{hR_{j,\xi,0}} \right] \frac{N_{k,j,\xi,.}}{N_{k,j,.,.}} \right\} \frac{N_{k,j,.,.}}{N_{k,.,.,.}}. \quad (14)$$

The Equation (14) is the lethality equation I introduced to compute the pandemic parameters of interest. We can notice that it depends on k only through the daily distribution of infection by age, VOC, and vaccination level and that we can derive the lethality by variant as

$$p_{k,\xi} = \sum_j \left(\sum_V RR_{j,\xi,V} \frac{N_{k,j,\xi,V}}{N_{k,j,\xi,.}} \right) \left[1 - \left(1 - p_{j,0,0} \right)^{hR_{j,\xi,0}} \right] \frac{N_{k,j,\xi,.}}{N_{k,.,\xi,.}},$$

where $\frac{N_{k,j,\xi,.}}{N_{k,.,\xi,.}}$ is the age distribution of infections due to the variant ξ. All still unknown quantities used to univocally determine the result will be specified in the paragraph Estimating the daily lethality.

Excess death

Let P and P_j be, respectively, the Italian population and its subgroup in the j-th age class and $Y_i^{(j)}$ be the binary random variable indicating that the ith person in the jth age class has been infected. If the virus spreads randomly within the population,

$$P\left\{ Y_i^{(j)} = 1 \right\} = \frac{1}{P} \forall\, i,j$$

we would have that the distribution of cases by age class equals that of the whole population

$$\sum_{i=1}^{P_j} P\left\{ Y_i^{(j)} = 1 \right\} = \frac{P_j}{P} \quad (15)$$

and the related probability of dying can be obtained from Equation (14) by replacing the proportion of infected people by age class ($\frac{N_{k,j,.,.}}{N_{k,.,.,.}}$) with the corresponding proportion in the population ($\frac{P_j}{P}$),

$$p_k^+ = \sum_j \left\{ \sum_\xi \left(\sum_V RR_{j,\xi,V} \frac{N_{k,j,\xi,V}}{N_{k,j,\xi,.}} \right) \left[1 - \left(1 - p_{j,0,0} \right)^{hR_{j,\xi,0}} \right] \frac{N_{k,j,\xi,.}}{N_{k,j,.,.}} \right\} \frac{P_j}{P}. \quad (16)$$

Since COVID-19 transmission began among younger people and eventually spread within the elderly (26), it was assumed that the spread was out-of-control if the distribution of detected cases by age class followed the age structure of the population (15). Through the deaths that resulted from the product between the Equation (2) and the Equation (16),

$$D_k^+ = \widehat{N}_k^{(D_k)} p_k^+ = \frac{\widehat{D}_k}{p_k} p_k^+,$$

the excess death was defined as the following difference:

$$D_k^{Excess} = D_k^+ - \widehat{D}_k.$$

Vaccine effect

Avoided infections

By rewriting the conditional probability in Equation (8) as ratio of probabilities, the relative risk in Equation (10) can be expressed as

$$RR_{j,\xi,V} = \frac{P\left\{ X_{j,\xi,V} = 1, Y_{k,j,\xi,V} = 1 \right\} / P\left\{ Y_{k,j,\xi,V} = 1 \right\}}{P\left\{ X_{j,\xi,0} = 1, Y_{k,j,\xi,0} = 1 \right\} / P\left\{ Y_{k,j,\xi,0} = 1 \right\}} \quad (17)$$

Under the assumption that the vaccines had no impact on the risk of catching the infection ($P\left\{ Y_{k,j,\xi,V} = 1 \right\} = P\left\{ Y_{k,j,\xi,0} = 1 \right\}$), the relative risk (17) reduces to:

$$RR_{j,\xi,V}^{(.,0)} = \frac{P\left\{ X_{k,j,\xi,V} = 1, Y_{k,j,\xi,V} = 1 \right\}}{P\left\{ X_{k,j,\xi,0} = 1, Y_{k,j,\xi,0} = 1 \right\}}. \quad (18)$$

Finally, let $Pop_{k,j,\xi,V}$ be the population at risk on the kth pandemic by age class, VOC, and vaccination level and $D_{k,j,\xi,V}$ be the number of deaths in each group, through the relationship

$$D_{k,j,\xi,V} = P\left\{ X_{j,\xi,V} = 1, Y_{k,j,\xi,V} = 1 \right\} Pop_{k,j,\xi,V},$$

the relative risk (18) can be rewritten as

$$RR_{j,\xi,V}^{(.,0)} = \frac{D_{k,j,\xi,V} / Pop_{k,j,\xi,V}}{D_{k,j,\xi,0} / Pop_{k,j,\xi,0}} \quad (19)$$

By replacing in Equation (14) $RR_{j,\xi,V}$ with $RR_{j,\xi,V}^{(.,0)}$, we obtain the daily probability $p_k^{(.,0)}$ to die if the vaccines have no protective effects against catching infection. By replacing in Equation (2) p_k with $p_k^{(.,0)}$, we can estimate the number of infections that would have occurred without vaccines

$$\widehat{N}_k^{(0)} = \frac{D_k}{p_k^{(.,0)}}. \quad (20)$$

Saved lives

If the vaccines have no effect against death, the relative risks $RR_{j,\xi,V}$ in Equation (14) would be equal to 1 and the lethality would reduce to

$$p_k^{(0,.)} = \sum_j \left\{ \sum_\xi \left[1 - \left(1 - p_{j,0,0} \right)^{hR_{j,\xi,0}} \right] \frac{N_{k,j,\xi,.}}{N_{k,j,.,.}} \right\} \frac{N_{k,j,.,.}}{N_{k,.,.,.}}. \quad (21)$$

By multiplying the Equation (20) for the Equation (21), we obtain an estimate of the number of deaths D_k^{**} that would have occurred without vaccines

$$D_k^{**} = \widehat{N}_k^{(0)} p_k^{(0,\cdot)}.$$

By multiplying the Equation (2) for the Equation (21), we obtain an estimate of the number of deaths D_k^{*} that would have occurred without vaccines among the infected people

$$D_k^{*} = \widehat{N}_k p_k^{(0,\cdot)}.$$

Checking estimates

We studied the ratios $\widehat{r}_k$ of detected cases $\widehat{\upsilon}_k$ among estimated infections $\widehat{N}_k$ on kth day (Figure 2)

$$\widehat{r}_k = \frac{\widehat{\upsilon}_k}{\widehat{N}_k}.$$

If $\widehat{r}_k\,(i) > 1$ (i.e., $\widehat{\upsilon}_k > \widehat{N}_k$), then the estimated $\hat{p}_k$ overestimates the actual p_k on the kth pandemic day

$$p_k < \hat{p}_k.$$

Estimating the daily lethality

Available data to estimate quantities in Equation (14) were used as follows:

1) The probability $p_{j,0,0}$ of dying among unvaccinated ($V = 0$) people in the jth age class who were infected with the original strain ($\xi = 0$) was estimated using the IFR by age class ($\widehat{IFR}_{j,0,0}$) in (6)

$$\hat{p}_{j,0,0} = \widehat{IFR}_{j,0,0}. \tag{22}$$

2) The daily unvaccinated population by age ($N_{k,j,.,0}$) was estimated as the difference between the ISTAT population[7] and the vaccinated people[4].
3) As estimates of hazard ratios $hR_{j,\xi,0}$ were used those from (27, 28) and since those for Alpha, Beta, Gamma, and Delta are not determined by age class, we considered them constant by age.
4) Let $D_{j,\xi,V}$, $C_{j,\xi,V}$, and $Pop_{j,\xi,V}$ be the number of deaths, of detected infections, and of population by age class, VOC, and vaccination level, respectively. The ISS provided estimates of the relative rate (vaccinated/unvaccinated) of deaths ($RD_{j,\xi,V}$) and of infections ($RC_{j,\xi,V}$)

$$\widehat{RD}_{j,\xi,V} = \frac{\widehat{D}_{j,\xi,V}/\widehat{Pop}_{j,\xi,V}}{\widehat{D}_{j,\xi,0}/\widehat{Pop}_{j,\xi,0}} \text{ and } \widehat{RC}_{j,\xi,V} = \frac{\frac{\widehat{C}_{j,\xi,V}}{\widehat{Pop}_{j,\xi,V}}}{\frac{\widehat{C}_{j,\xi,0}}{\widehat{Pop}_{j,\xi,0}}},$$

for the periods January–September/2021, October/2021, November/2021, December/2022, January /2022, and February/2022 (22). We used those estimates to assess the relative risk in Equation (10) and in Equation (19) as follows

$$\widehat{RR}_{j,\xi,V} = \frac{\widehat{RD}_{j,\xi,V}}{\widehat{RC}_{j,\xi,V}} = \frac{\widehat{D}_{j,\xi,V}/\widehat{C}_{j,\xi,V}}{\widehat{D}_{j,\xi,0}/\widehat{C}_{j,\xi,0}} \text{ and}$$
$$\widehat{RR}_{j,\xi,V}^{(.,0)} = \frac{\widehat{RD}_{j,\xi,V}}{\widehat{RC}_{j,\xi,0}} = \widehat{RD}_{j,\xi,V}. \tag{23}$$

5) As estimates of the fraction of vaccinated infections by age class and VOC ($\frac{N_{k,j,\xi,V}}{N_{k,j,\xi,.}}$), the daily fraction of vaccinated population by age class[4] were used. The more a VOC is prevalent, the lesser the introduced bias.
6) As daily fraction of infections for each VOC ($\xi = 0, 1, \ldots, 5$) by age class ($\frac{N_{k,j,\xi,.}}{N_{k,j,.,.}}$) daily estimates from[5] were used.
7) As estimates of age distribution of total infections ($\frac{N_{k,j,.,.}}{N_{k,.,.,.}}$) and of those by VOC ($\frac{N_{k,j,\xi,.}}{N_{k,.,\xi,.}}$), was used the daily age distribution of detected cases released by the ISS from December 8, 2020 ($f_{k,j}^{(ISS)}$)[3]. For the precedent period (during which the ISS only released the median age of detected cases), we constructed fictitious populations $P^{(Med_k)}$ with median ages (MEDs) equal to those estimated and with the age structure related with that of Italian population provided by the ISTAT[7] [Section Estimating $f_{j,k}^{(Med_k)}$].

Estimating $f_{j,k}^{(Med_k)}$

Let P_j and $P_{k,j}^{(Med_k)}$ be the people of age j in the Italian and in fictitious populations, respectively, on the kth pandemic day, using the definition of "median" we have that

$$P_{k,0}^{(Med_k)} \sum_{j=0}^{Med_k} \frac{P_{k,j}^{(Med_k)}}{P_{k,0}^{(Med_k)}} = 0.5 \text{ and}$$
$$P_{k,100+}^{(Med_k)} \sum_{j=Med_k+1}^{100} \frac{P_{k,j}^{(Med_k)}}{P_{k,100+}^{(Med_k)}} = 0.5, \tag{24}$$

where 100+ indicate people aged 100 years and more. By assuming that the ratios between infected people at ages greater than the median and the oldest (100+ years) are equal to those in the Italian population, and that the ratios between those at ages smaller than or equal to the median

and the youngest (0 years) are also equal to those in the Italian population

$$\frac{P_{k,j}^{(Med_k)}}{P_{k,0}^{(Med_k)}} = \frac{P_j}{P_0} \text{with } j = 0, 1, \ldots, MED \text{ and}$$
$$\frac{P_{k,j}^{(Med_k)}}{P_{k,100+}^{(Med_k)}} = \frac{P_j}{P_{100+}} \text{with } j = MED{+}1, \ldots, 100{+}, \quad (25)$$

$P_{k,0}^{(Med_k)}$ and $P_{k,100+}^{(Med_k)}$ are determined from the Equation (24) and can be used to derive all the remaining fractions ($f_{k,j}^{(Med_k)}$)

$$f_{k,j}^{(Med_k)} = \frac{P_{k,j}^{(Med_k)}}{P_{k,.}^{(Med_k)}} \quad j = 1, \ldots, 100+.$$

Since the method returns one probability estimate per week (ISS median age refers to a week), each pair of values was linearly connected. By assuming $RR_{j,\xi,V} = hR_{j,\xi,0} = 1$ in Equation (14), the resulting death probability is equal to

$$p_k^{(Med_k)} = \sum_j p_{j,0,0} f_{k,j}^{(Med_k)}. \quad (26)$$

By replacing the Equation (26) in the Equation (2), we estimated the number of infections ($\widehat{N}$) from the beginning to the middle day of the ISTAT seroruvey (June 19, 2020; after 120 gg from the beginning) as

$$\widehat{N} = \sum_{k=1}^{120} \frac{D_k}{p_k^{(Med_k)}}$$

and the ratio with the corresponding ISTAT estimate ($\overline{N}$) was used as correction factor of the Equation (26)

$$\hat{p}_k = \frac{\widehat{N}}{\overline{N}} p_k^{(Med_k)} \text{with } k = 1, ..., 120. \quad (27)$$

Waves

In epidemiology, an internationally accepted definition of "wave" does not still exist; the term refers to the appearance of a plot of cases over time. In this paper, the word "wave" is used to indicate the part of the plot that lies between two local minima.

Health policy evaluation

The effects of applied health policies were determined by comparing the weekly and bi-weekly incidence rates before and after the day (k) they entered into force

$$\frac{\left(N_{k+j} - N_k\right) - \left(N_k - N_{k-j}\right)}{N_k - N_{k-j}} \text{ with } k \in Z^+ \text{ and } j = 7, 14.$$

Data cleaning methods

Dates of the ISS cumulative distribution of detected cases by age correlate with the most recent update. Those distributions were reordered to have non-decreasing functions. Since the VOC prevalence data in TESSEy cover a week, the data were linearly fitted to obtain daily prevalence. In addition, frequencies recorded before the official date of the first available samples were set to zero.

Results

During the first 2 years of the COVID-19 pandemic (from February 24, 2020, to February 28, 2022; 736 days and 106 weeks) in Italy, there were 20,833,018 (20,728,922–20,937,373) infections and 152,358 deaths for an overall lethality rate of 0.73%. Health Institutions detected 63% of the total infections using 193,442,203 tests. From February 1, 2021, three VOCs became predominant: Alpha (February–June, 2021); Delta (July–December, 2021); and Omicron (January–February, 2022). Up to February 2022, 14.2% of people were unvaccinated (including children aged 0–4 years), 2.1% were waiting for the second shot, 12.2% had completed the basic vaccine cycle (one shot for the Jensen vaccine or two shots of other vaccines) > 4 months ago, 7.9% had completed the basic vaccine cycle in the last 4 months, and 63.6% had received the additional dose. Health policies evolved from a national lockdown to the development of a strong contact tracing system (the monthly number of tests increased from 506,496 in March 2020 to 17,089,550 in February 2022), accompanied by a large-scale vaccination campaign and additional local measures.

Incidence curve

The incidence curve in Figure 1 shows five waves. The first wave (characterized by infection with non-VOCs variants) lasted 147 days (from February 24 to July 19, 2020), included 1,526,561 infections (of which 240,802 were detected through 6,937,326 diagnostic tests), and peaked at 33,683 infections on March 12, 2020. The second wave (characterized by the predominance of the Alpha variant in the right tail) lasted 200 days (from July 20, 2020, to February 4, 2021), included 4,716,509 infections (of which 2,495,649 were detected through 29,891,933 diagnostic tests), and peaked at 57,594 infections on November 3, 2020. This wave also showed a hump of about 30,000 infections in the second half of December. The third wave (characterized by the Alpha variant) lasted 139 days (from February 5 to June 23, 2021), included 2,800,141 infections (of which 1,553,334 were detected through 37,736,565 diagnostic tests), and peaked at 36,471 infections on March 11, 2021. The fourth wave (characterized by the Delta variant) lasted 100 days

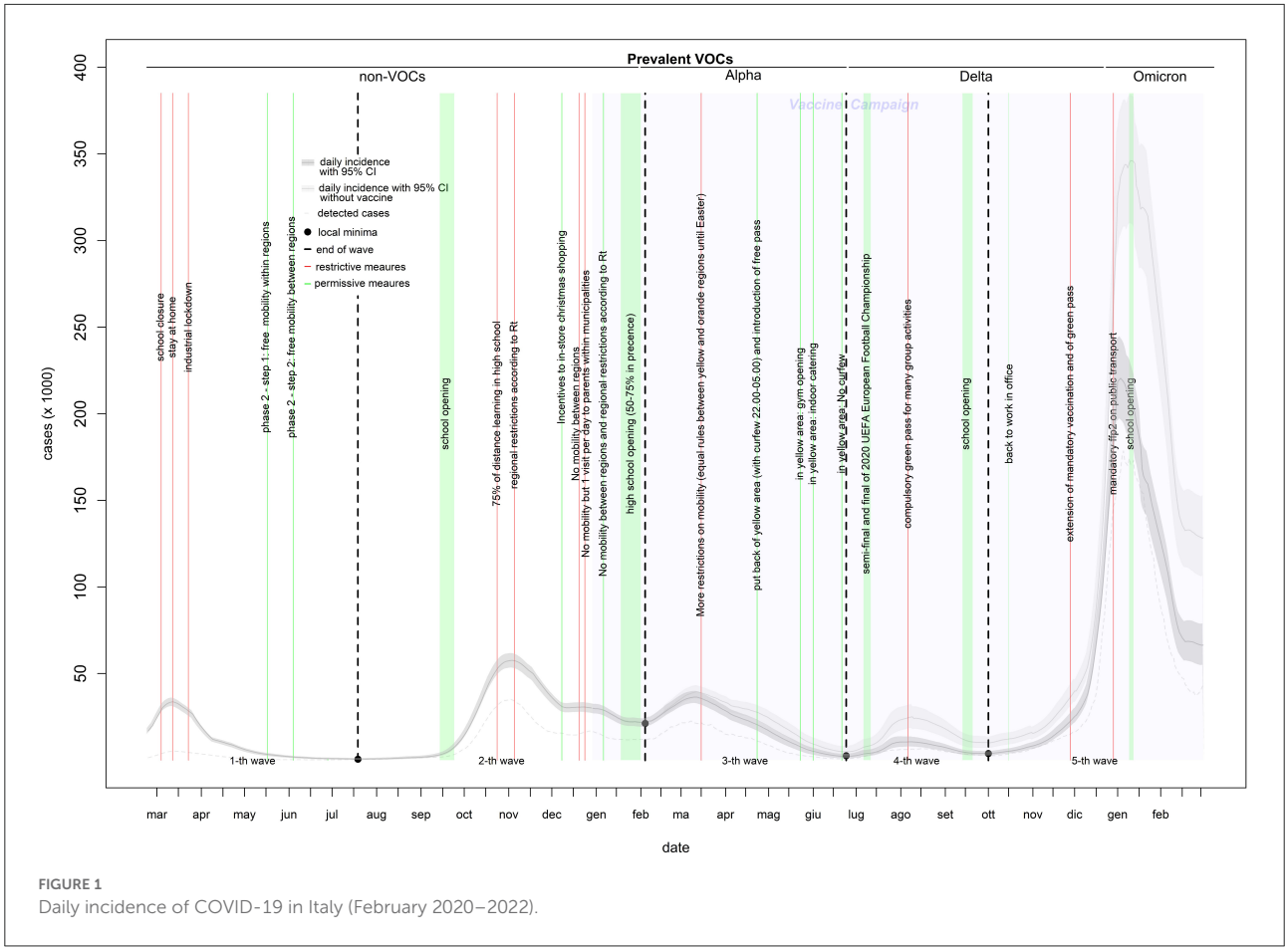


FIGURE 1
Daily incidence of COVID-19 in Italy (February 2020–2022).

(from June 24, 2021, to October 1, 2021), included 650,487 infections (of which 436,315 were detected through 23,693,891 diagnostic tests), and peaked at 10,617 infections on August 9, 2021. The fifth wave (characterized by a mixture of the Delta and Omicron variants) lasted 150 days (from October 2, 2021, to the end of the study period), included 11,139,320 infections (of which 8,344,165 were detected through 98,182,488 diagnostic tests), and peaked at 220,487 infections on January 1, 2022 (Table 1). From the third to the fifth wave, the proportion of infections among young individuals (0–19 years) increased by up to 30% (Supplementary Figure 1).

Lethality

Estimates of infection-related deaths during the 1st months of the pandemic were provided by a fictitious population (26) and adjusted using a correction factor (27) of 1.02%. Daily lethality ranged from 2.8% (first wave: April 9, 2020) to 0.15% (last wave: December 30, 2021), causing a total of 152,358 deaths with a peak of 723 deaths on November 5, 2020 (Figures 2, 3). After the peak, lethality in the first wave decreased to 1.0% on June 11, 2020, generating 32,739 deaths. Initially, lethality of the second wave, which caused 62,595 deaths, decreased to 0.8% on August 15, 2020, and then increased to a peak of 1.6% on December 8, 2020, then remained stable. The lethality in the third wave peaked at 1.4% on February 8, 2021, then decreased to 0.45% by June 8, 2021, and caused 28,596 deaths. The lethality in the fourth wave, which caused 3,620 deaths, initially decreased to 0.38% by July 6, 2021, and then increased to a peak of 0.91% by September 30, 2021. The lethality of the fifth wave, which caused 24,808 deaths, decreased continuously from a peak of 0.91% on October 2, 2021 to 0.15% on December 30, 2021, and then slowly increased to 0.18% by the end of February 2022. Two periods had the lethality higher than the threshold: the first 4 months of the first wave (March-June, 2020), with an excess of 14,134 deaths; and the last 3 months of the second wave (November, 2020–January, 2021), with an excess of 6,478 deaths (Figure 3).

Impact of variants of concerns

Of five VOCs, three became prevalent (Alpha, Delta, and Omicron), each replacing prior variants at a faster

TABLE 1 Observed vs. expected (without vaccines) epidemiology of COVID-19 pandemic in Italy (February 2020–2022).

Wave	Period	Observed pandemic					Expected pandemic without vaccine*			Diagnosis	Test
		Deaths	Infection (95% CI)	Lethality (%)	Max**	Day of max**	Infection (95% CI)	Deaths	Lethality (%)		
1	24/02/2020 19/07/2020	32,739	1,526,561 (1,542,962–1,510,247)	2.14	33,683	12/03/2020	1,526,561 (1,542,962–1,510,247)	32,739	2.14	240,802	6,937,326
2	20/07/2020 04/02/2021	62,595	4,716,509 (4,753,282–4679877)	1.33	57,594	03/11/2020	4,722,972 (4,759,726–4,686,359)	62,830	1.33	2,495,649	29,891,933
3	05/02/2021 23/06/2021	28,596	2,800,141 (2,832,521–2,767,945)	1.02	36,471	11/03/2021	3,369,666 (3,402,224–3,337,263)	40,844	1.21	1,553,334	37,736,565
4	24/06/2021 01/10/2021	3,620	650,487 (671,789–629,524)	0.56	10,617	09/08/2021	1,519,344 (1,540,474–1,498,358)	19,737	1.30	436,315	23,693,891
5	02/10/2021 28/02/2022	24,808	11,139,320 (11,278,202–11,001,288)	0.22	220,487	01/01/2022	19,135,912 (19,248,680–19,023,470)	110,298	0.58	8,344,165	98,182,488
Total	24/02/2020 28/02/2022	152,358	20,833,018 (20,728,922–20,937,373)	0.73	220,487	01/01/2022	30,274,455 (30,160,115–30,389,006)	266,448	0.88	13,070,266	196,442,203

*Vaccine campaign started at the end of the second wave (2020/12/27).
**Max = max number of daily infections.

pace. Omicron was responsible for the higher number of infections (7,557,368), while Alpha was associated with the largest number of deaths (29,167). Without vaccines, Delta would be the most virulent VOC with > 70,000 deaths (Supplementary Table 1).

Vaccine effect

Vaccines reduced infections by 38% (from 25,045,987 to 15,604,551) and deaths by 62% (from 185,850 to 71,760). Of 114,090 lives saved, 62,902 (55%) would have resulted from infections prevented and 51,188 (45%) from the infections that occurred. Without vaccines, the expected number of infections would have been 30,274,455 (30,160,115–30,389,006), and the expected number of deaths 266,448 for a lethality of 0.88% (Figures 1–3 and Table 1).

Health policies effects on estimated curves

The strongest restriction measures affecting all the population (initial stay at home, the November 2020 introduction of standardized prevention measures based geographic risk, and the Christmas 2020 and Easter 2021 restrictions) strongly reduced the curve rates (up to 1,000%) within a 1st week of their introduction. Industrial lockdown and specific restrictions (including 75% of high-school students who received distance learning) implemented in October 2020 are associated with smaller (from −37 to −227%) and slower (concentrated in the 2nd week) rate reductions. Curve rates increased after school openings (except after the last one of January 2022) and reductions in smart working, especially during the 2nd week. Rate increments after school openings reduced over time. In-shop Christmas 2020 incentives increased the incidence rate by 90–100% and the death rate by 45–60%. The introduction of a compulsory green pass reduced the rate of infection and death curve by 150–200 and 40–70%, respectively. The mandatory use of the FFp2 mask in closed places reduced the curve rates by 45–85% (Table 2). The introduction of rapid tests (from January 2021) increased the percentage of infections detected among children, particularly when schools were open (Supplementary Figure 1).

Discussion

This paper provides a comprehensive picture for the first 2 years (February 2020–2022) of the COVID-19 pandemic in Italy, including the impact of VOCs, the vaccine campaign (until the third shot), and an evaluation of government health policies using only public data.

TABLE 2 Differences in rates of COVID-19 incidence and death curves before and after prevention measures.

N	Date	Measures	Relative difference of rates (%)			
			Incidence curve		death curve	
			Weekly	Bi-weekly	Weekly	Bi-weekly
1	05/03/2020	Schools closed	−67	−98	−73	−105
2	12/03/2020	Stop to all mobility—Stay at home	−189	−152	−238	−179
3	23/03/2020	Industrial lockdown	−77	−227	−77	−86
4	17/05/2020	Allowing intraregional mobility	30	48	35	57
5	04/06/2020	Free mobility	29	40	10	27
6	14/09/2020	Schools opened	609	1,128	544	1,027
7	24/10/2020	Several restrictions (including 75% DAD high school)	−57	−83	−37	−63
8	05/11/2020	Regional restrictions according to Rt	−508	−220	−151	−149
9	08/12/2020	Incentives for Christmas shopping	90	97	46	59
10	20/12/2020	No mobility between regions	−179	37	32	9
11	24/12/2020	Christmas rules: Just 1 visit per day to parents within municipalities	−643	−1,002	−62	−43
12	07/01/2021	Regional restrictions according to Rt	−182	−184	−33	−77
13	25/01/2021	High school opening (50–75% in presence)	124	175	123	150
14	15/03/2021	Easter rules: equal rules between yellow and orange regions	−458	−433	−239	−2,497
15	23/04/2021	Put back of yellow are (with curfew 22.00–05.00) and introduction of free pass	−21	−39	12	23
16	23/05/2021	In yellow area: Gym opening	15	33	28	50
17	01/06/2021	In yellow area: Indoor catering	29	47	33	56
18	21/06/2021	In yellow area: No curfew	118	167	94	133
19	09/07/2021	Semi-final and final of 2020 UEFA European Football Championship	38	257	188	497
20	06/08/2021	Compulsory green pass for many group activities	−208	−147	−41	−71
21	17/09/2021	Schools opened	90	97	94	120
22	15/10/2021	Back to work in office	24	85	27	66
23	27/11/2021	Extension of mandatory vaccination and of green pass	32	135	24	63
24	27/12/2021	Mandatory ffp2 on public transport	−55	−85	−44	−76
25	09/01/2022	Schools opened	−180	−1,524	−2,842	−309

Italy, February 2020–2022.

Virus spread

Almost 40% of COVID-19 infections have gone undetected, likely because they were asymptomatic or paucisymptomatic. During the first wave, the virus primarily spread in the north of the country and was highly concentrated in the Lombardy region. The virus likely arrived in Italy through the airport system of Milan (the largest city in Lombardy), which includes one intercontinental and two international airports (one of which in Bergamo, the most hard-hit Italian city). Like other respiratory viruses, SARS-CoV-2 spreads directly or indirectly through person-to-person contact (especially in indoor environments) (29). Lombardy is the Italian region with the highest level of daily commuting for work or school[8]. Indeed, a study highlighted the association between the regional patterns of viral spread during the first wave and the origin-destination matrix of goods and food transportation and for the population (30). Another component responsible for the rapid spread of COVID-19 was unpreparedness. During the onset phase of the pandemic, hospitals followed WHO guidelines and tested only people with a known link to China (thus accelerating the spread of the virus). Fortunately, the quick stop to all national mobility on March 12, 2020, confined the virus to northern regions. Data collected during the second wave revealed similar patterns to those reported by other studies: the virus infects younger people first followed by those >70 years of age (26). Since retired people have fewer daily contacts than students and workers, who often use public transport and share indoor environments with others, it is likely that school and work transmission impacted the onset of the familial transmission chain among the elderly. Infected students and workers carried the virus home, transmitted the infection to other family members (31) and increased the probability that older relatives (including grandparents) would become infected, especially through presymptomatic or asymptomatic infections. During the summer months, those of 20–29 years of age were the

8 https://www.istat.it/it/archivio/139381

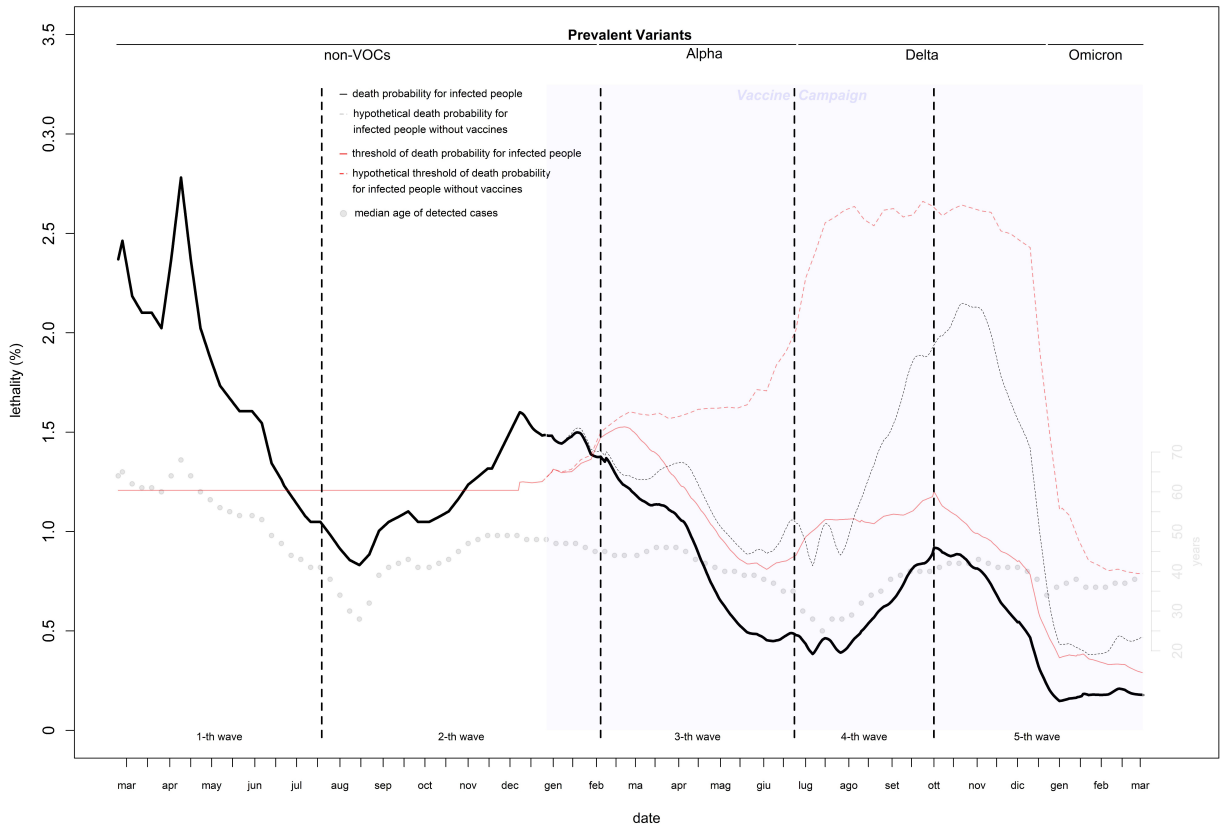


FIGURE 2
Daily lethality of COVID-19 in Italy (February 2020–2022).

hardest hit, presumably because of increased nightlife and other social activities. Without public health policies, it is likely that two waves would have occurred per year (similar to the fourth and fifth waves as shown in Figure 1): a winter wave (resulting from a higher number of transmissions from indirect contacts) and a summer wave (largely resulting from direct contacts). The former is longer (October to June), peaks in January, and is more virulent because it mainly involves families; the latter is shorter (July to September), peaks in August, and is less virulent because it primarily involves single people.

Lethality

The estimates of lethality obtained by the fictitious populations (26) provided a cases count by June 20, 2020, that was very close to that estimated by the national ISTAT sero-survive[6], with an error of 2%. This indicates that with a high median age of detected cases (with respect to that of the national population), the fictitious population provides reasonable estimates when the age distribution of cases is unknown. The virus's lethality was extremely high in the first 3 months of the pandemic, when the median age of detected cases was much higher than that of the population. This is the result of two serious errors: a lack of screening tests to reduce transmission from adult/young to elderly and the use of nursing homes to support hospitals with saturated capacity in the hardest hit regions (which increased infections among the most at risk population) (32). Lethality was higher than expected even during the second wave. The introduction of reliable rapid tests allowed a massive test campaign that kept lethality under the threshold from the right tail of the second wave. Lethality would have been under the threshold even without vaccines (Figure 2 dotted curves). Of the three prevalent VOCs in Italy, the first two were more virulent than the original strain, the third was not. To ensure that vaccines remain updated and new and potentially harmful variants are identified early, ongoing research on the evolution of virus genomes is crucial.

Health policies

The stronger the restriction measure, the higher its efficacy and the more quickly it took effect. Supporting the assumption that the incidence curve was largely underestimated at the onset of pandemic, the initial lockdown impacted the death rate more than the infections rate. After the second wave, the large-scale screening helped to monitor the actual size

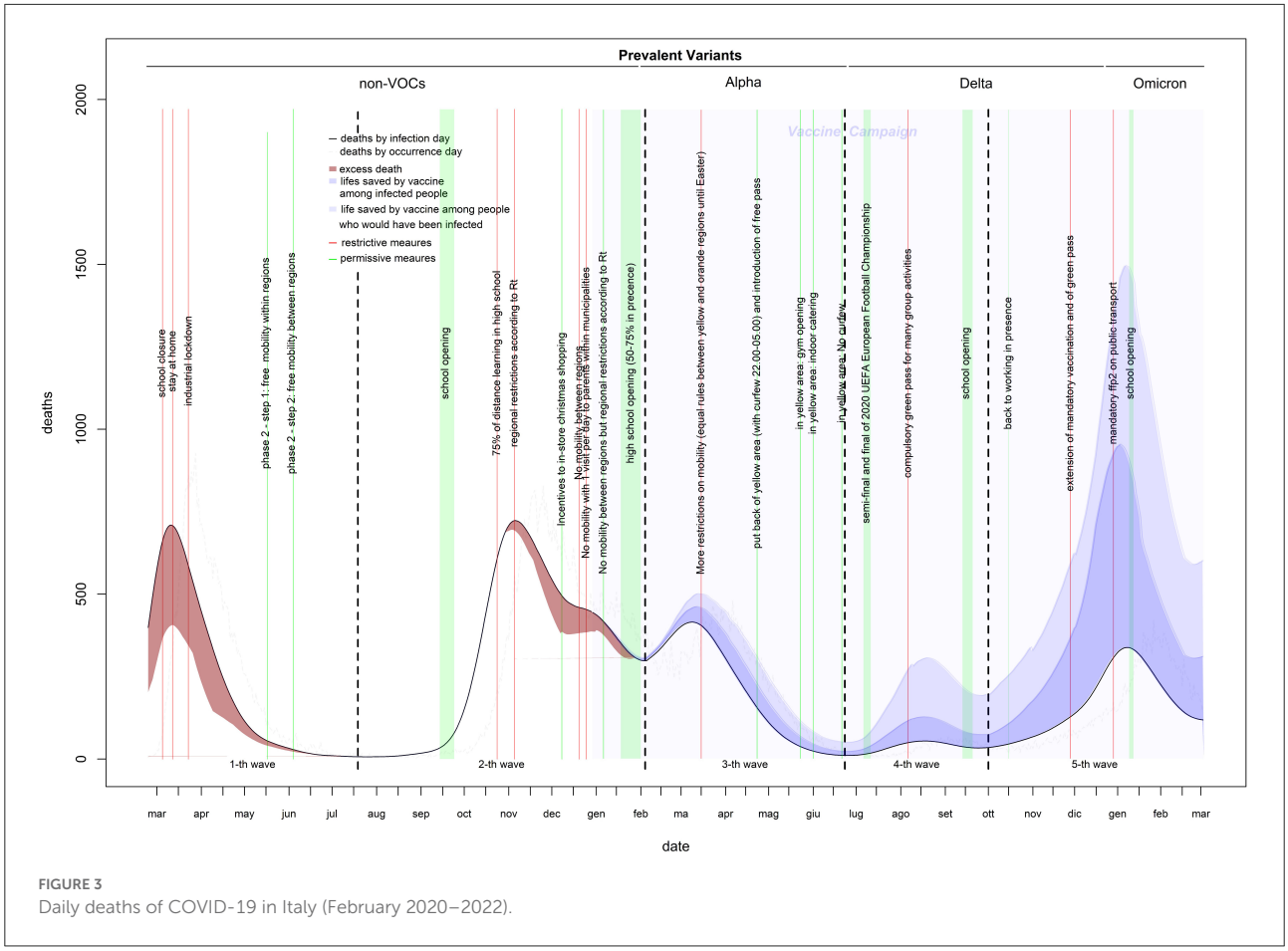


FIGURE 3
Daily deaths of COVID-19 in Italy (February 2020–2022).

of the outbreaks, especially among young students (often asymptomatic). Quarantining infected grandsons (tested at school) and parents (tested at work) likely protected the grandparents. This is supported by increased rates of infection and death curves after school openings (except the last one) and reductions in smart working. However, in schools, those increments declined over time until disappearing, presumably because school protocols became more and more effective. Even if the mandatory use of the ffp2 in the public transport reduced the curves rate by up to 70% during 2021 Christmas holidays, their true effect is shown after the schools opening, where rates drastically decreased of 1,500–2,800%. Although protective effect of the vaccines was reduced by the emergence of new variants, vaccination saved more than 110,000 lives and avoided the saturation of the health system without a need for stronger restriction measures even during periods of high virus circulation.

Advice

It is necessary to monitor the evolution of SARS-CoV-2 in greater depth and to develop mathematical models that can predict future changes in its genome. A flexible pandemic plan able to adapt to the evidence of the data (collected through a digital and multi-connected surveillance system) should be developed through a multidisciplinary approach and shared with international health authorities. It should contain measures that are tailored for different combinations of virus transmissibility and virulence (from low-low to high-high). Estimates of territorial origin-destination matrices can help to simulate possible spatiotemporal patterns of virus spread. Initial settings of a public health response should refer to an "average" or "worst case" scenario and updates should follow data evidence.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://github.com/floatingpurr/covid-19_sorveglianza_integrata_italia/tree/main/data.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with

the local legislation and institutional requirements. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

PF: study concept and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content, and statistical analysis.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.986743/full#supplementary-material

References

1. Klein S, Cortese M, Winter SL, Wachsmuth-Melm M, Neufeldt CJ, Cerikan B, et al. SARS-CoV-2 structure and replication characterized by *in situ* cryo-electron tomography. *Nat Commun.* (2020) 11:1–10. doi: 10.1038/s41467-020-19619-7

2. Barth RF, Buja LM, Parwani AV. The spectrum of pathological findings in coronavirus disease (COVID-19) and the pathogenesis of SARS-CoV-2. *Diagn Pathol.* (2020) 15:1–4. doi: 10.1186/s13000-020-00999-9

3. Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. *Acta Bio Medica*. (2020) 91:157. doi: 10.23750/abm.v91i1.9397

4. Wei WE Li Z, Chiew CJ, Yong SE, Toh MP, Lee VJ. Presymptomatic transmission of SARS-CoV-2 — Singapore, January 23–March 16, 2020. *Morb Mortal Wkly Rep.* (2020) 69:411–5. doi: 10.15585/mmwr.mm6914e1

5. Ye F, Xu S, Rong Z, Xu R, Liu X, Deng P, et al. Delivery of infection from asymptomatic carriers of COVID-19 in a familial cluster. *Int J Infect Dis.* (2020) 94:133–8. doi: 10.1016/j.ijid.2020.03.042

6. Brazeau N, Verity R, Jenks S, Fu H, Whittaker C, Winskill P, et al. *Report 34: COVID-19 Infection Fatality Ratio: Estimates From Seroprevalence.*

7. O'Driscoll M, Ribeiro Dos Santos G, Wang L, Cummings DA, Azman AS, Paireau J, et al. Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature.* (2021) 590:140–5. doi: 10.1038/s41586-020-2918-0

8. Ghisolfi S, Almås I, Sandefur JC, von Carnap T, Heitner J, Bold T. Predicted COVID-19 fatality rates based on age, sex, comorbidities and health system capacity. *BMJ Glob Health.* (2020) 5:e003094. doi: 10.1136/bmjgh-2020-003094

9. Vaid N, Ardissino M, Reed TA, Goodall J, Utting P, Miscampbell M, et al. Clinical characteristics and outcomes of immunosuppressed patients hospitalized with COVID-19: experience from London. *J Internal Med.* (2021) 289:385–94. doi: 10.1111/joim.13172

10. Perone G. The determinants of COVID-19 case fatality rate (CFR) in the Italian regions and provinces: an analysis of environmental, demographic, and healthcare factors. *Sci Tot Environ.* (2021) 755:142523. doi: 10.1016/j.scitotenv.2020.142523

11. Li Y, Tenchov R, Smoot J, Liu C, Watkins S, Zhou Q. A comprehensive review of the global efforts on COVID-19 vaccine development. *ACS Central Sci.* (2021) 7:512–33. doi: 10.1021/acscentsci.1c00120

12. Francis AI, Ghany S, Gilkes T, Umakanthan S. Review of COVID-19 vaccine subtypes, efficacy and geographical distributions. *Postgrad Medical J.* (2022) 98:389–94. doi: 10.1136/postgradmedj-2021-140654

13. Castonguay N, Zhang W, Langlois MA. Meta-analysis and structural dynamics of the emergence of genetic variants of SARS-CoV-2. *Front Microbiol.* (2021) 12:1637. doi: 10.3389/fmicb.2021.676314

14. Hadj Hassine I. Covid-19 vaccines and variants of concern: a review. *Rev Medical Virol.* (2022) 32:e2313. doi: 10.1002/rmv.2313

15. Yang Z, Zhang S, Tang YP, Zhang S, Xu DQ, Yue SJ, et al. Clinical characteristics, transmissibility, pathogenicity, susceptible populations, and re-infectivity of prominent COVID-19 variants. *Aging Dis.* (2022) 13:402–22. doi: 10.14336/AD.2022.0307

16. McIntyre PB, Aggarwal R, Jani I, Jawad J, Kochhar S, MacDonald N, et al. COVID-19 vaccine strategies must focus on severe disease and global equity. *Lancet.* (2022) 399:406. doi: 10.1016/S0140-6736(21)02835-X

17. Tartof SY, Slezak JM, Fischer H, Hong V, Ackerson BK, Ranasinghe ON, et al. Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: a retrospective cohort study. *Lancet.* (2021) 398:1407–16. doi: 10.1016/S0140-6736(21)02183-8

18. Bar-On YM, Goldberg Y, Mandel M, Bodenheimer O, Freedman L, Kalkstein N, et al. Protection of BNT162b2 vaccine booster against Covid-19 in Israel. *N Engl J Med.* (2021) 385:1393–400. doi: 10.1056/NEJMoa2114255

19. Camilla M, Lippi G. Primary COVID-19 vaccine cycle and booster doses efficacy: analysis of Italian nationwide vaccination campaign. *Eur J Public Health.* (2022) 2022:ckab220. doi: 10.1093/eurpub/ckab220

20. Miraglia del Giudice G, Napoli A, Corea F, Folcarelli L, Angelillo IF. Evaluating COVID-19 vaccine willingness and hesitancy among parents of children aged 5–11 years with chronic conditions in Italy. *Vaccines.* (2022) 10:396. doi: 10.3390/vaccines10030396

21. Ferrante P. The first year of COVID-19 in Italy: incidence, lethality, and health policies. *J Public Health Res.* (2022) 11:jphr-2021. doi: 10.4081/jphr.2021.2201

22. Istituto Superiore di Sanità. *Epidemia COVID-19. Aggiornamento nazionale: 16 febbraio 2022.* (2022). https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19_30-marzo-2022.pdf (last accessed 09/19/2022)

23. De Natale G, Ricciardi V, De Luca G, De Natale D, Di Meglio G, Ferragamo A, et al. The COVID-19 infection in Italy: a statistical study of an abnormally severe disease. *J Clin Med.* (2020) 9:1564. doi: 10.3390/jcm9051564

24. Istituto Superiore di Sanità. *Epidemia COVID-19. Aggiornamento nazionale: 18 agosto 2020.* (2020). Available online at: https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19_18-agosto-2020.pdf (accessed September 09, 2022).

25. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported

confirmed cases: estimation and application. *Ann Intern Med.* (2020) 172:577–82. doi: 10.7326/M20-0504

26. Tran Kiem C, Bosetti P, Paireau J, Crepey P, Salje H, Lefrancq N, et al. SARS-CoV-2 transmission across age groups in France and implications for control. *Nat Commun.* (2021) 12:1–12. doi: 10.1038/s41467-021-27163-1

27. Li L, Liu Y, Tang X, He D. The disease severity and clinical outcomes of the SARS-CoV-2 variants of concern. *Front Public Health.* (2021) 2021:1929. doi: 10.3389/fpubh.2021.775224

28. Nyberg T, Ferguson NM, Nash SG, Webster HH, Flaxman S, Andrews N, et al. Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 Omicron (B. 11 529) and Delta (B 1617 2) variants in England. *Lancet.* (2022) 399:1303–12. doi: 10.1016/S0140-6736(22)00462-7

29. Boncristiani HF, Criado MF, Arruda E. Respiratory viruses. *Encycl Microbiol.* (2009) 500:314. doi: 10.1016/B978-012373944-5.00314-X

30. Tosi D, Chiappa M. Understanding the geographical spread of COVID-19 in relation with goods regional routes and governmental decrees: the Lombardy Region Case Study. *SN Comput Sci.* (2021) 2:1–9. doi: 10.1007/s42979-021-00597-6

31. Haroon S, Chandan JS, Middleton J, Cheng KK. Covid-19: breaking the chain of household transmission. *Br Med J.* (2020) 370. doi: 10.1136/bmj.m3181

32. Logar S. Care home facilities as new COVID-19 hotspots: Lombardy Region (Italy) case study. *Archiv Gerontol Geriatr.* (2020) 89:104087. doi: 10.1016/j.archger.2020.104087

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 14 November 2022
DOI 10.3389/fpubh.2022.992697

OPEN ACCESS

EDITED BY
Pierpaolo Ferrante,
National Institute for Insurance Against Accidents at Work (INAIL), Italy

REVIEWED BY
Hongsong Chen,
Peking University People's Hospital, China
Jun Tanimoto,
Kyushu University, Japan

*CORRESPONDENCE
Hsiang-Yu Yuan
sean.yuan@cityu.edu.hk

SPECIALTY SECTION
This article was submitted to Infectious Diseases - Surveillance, Prevention and Treatment, a section of the journal Frontiers in Public Health

RECEIVED 12 July 2022
ACCEPTED 28 September 2022
PUBLISHED 14 November 2022

CITATION
Lin C-P, Dorigatti I, Tsui K-L, Xie M, Ling M-H and Yuan H-Y (2022) Impact of early phase COVID-19 precautionary behaviors on seasonal influenza in Hong Kong: A time-series modeling approach.
Front. Public Health 10:992697.
doi: 10.3389/fpubh.2022.992697

Impact of early phase COVID-19 precautionary behaviors on seasonal influenza in Hong Kong: A time-series modeling approach

Chun-Pang Lin[1,2], Ilaria Dorigatti[3], Kwok-Leung Tsui[4], Min Xie[1], Man-Ho Ling[5] and Hsiang-Yu Yuan[6]*

[1]School of Data Science, City University of Hong Kong, Kowloon, Hong Kong SAR, China, [2]Department of Statistics, School of Arts and Sciences, Rutgers University, New Brunswick, NJ, United States, [3]MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London, United Kingdom, [4]Grado Department of Industrial and Systems Engineering, College of Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States, [5]Department of Mathematics and Information Technology, Faculty of Liberal Arts and Social Sciences, The Education University of Hong Kong, Tai Po, Hong Kong SAR, China, [6]Department of Biomedical Sciences, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Kowloon, Hong Kong SAR, China

Background: Before major non-pharmaceutical interventions were implemented, seasonal incidence of influenza in Hong Kong showed a rapid and unexpected reduction immediately following the early spread of COVID-19 in mainland China in January 2020. This decline was presumably associated with precautionary behavioral changes (e.g., wearing face masks and avoiding crowded places). Knowing their effectiveness on the transmissibility of seasonal influenza can inform future influenza prevention strategies.

Methods: We estimated the effective reproduction number (R_t) of seasonal influenza in 2019/20 winter using a time-series susceptible-infectious-recovered (TS-SIR) model with a Bayesian inference by integrated nested Laplace approximation (INLA). After taking account of changes in underreporting and herd immunity, the individual effects of the behavioral changes were quantified.

Findings: The model-estimated mean R_t reduced from 1.29 (95%CI, 1.27–1.32) to 0.73 (95%CI, 0.73–0.74) after the COVID-19 community spread began. Wearing face masks protected 17.4% of people (95%CI, 16.3–18.3%) from infections, having about half of the effect as avoiding crowded places (44.1%, 95%CI, 43.5–44.7%). Within the current model, if more than 85% of people had adopted both behaviors, the initial R_t could have been less than 1.

Conclusion: Our model results indicate that wearing face masks and avoiding crowded places could have potentially significant suppressive impacts on influenza.

KEYWORDS

COVID-19, influenza, face mask, social distancing, time-series analysis, infectious disease modeling

1. Introduction

Many studies warned that repeated COVID-19 outbreaks are expected to happen and the number of infections and deaths could become even worse during winter (1–5). Besides the relaxation of social distancing during winter holidays, seasonal influenza, which commonly circulates during wintertime, may facilitate the transmission and mortality of COVID-19 if both severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and influenza virus spread at the same time (6–9). In fact, many places have seen surges in COVID-19 infections in 2020/2021 winter (10–12). Since many cities have reopened after the vaccine has been distributed, it is important to know whether individual precautionary behaviors (such as avoiding crowded places and wearing face masks) without strict social distancing rules can prevent an influenza outbreak. How to prevent influenza outbreak or the co-circulation with COVID-19 are major tasks for World Health Organization (13, 14).

During the early spread of COVID-19 in China, many Hong Kong residents began to wear face masks mainly in public transport and avoid going to crowded places voluntarily. Due to the high influx of travelers from mainland China, Hong Kong faced and acknowledged an extremely high risk during the early spread. After WHO made an announcement of the initial spreading of COVID-19 in Wuhan on January 14, 2020 (15), people in Hong Kong perceived the risk of infection and changed their behavior immediately. Cowling et al. (16) showed that the number of people avoiding crowded places and wearing face masks increased between January and February 2020. Their first survey (January 20–23) was conducted immediately after the announcement by WHO about noting limited human-to-human transmission and the First-Level Public Health Emergency Response in China (15, 17). Their second survey (February 11–14) was conducted after the first local (community) transmission event was confirmed in Hong Kong on February 4. Few cases sporadically occurred up to early March, indicating that the first significant COVID-19 outbreak began. Because seasonal influenza incidence was progressively reducing during these survey periods soon after its initial rapid growth, it is likely that this unexpected reduction in the incidence of influenza was due to the behavioral changes in response to the potential COVID-19 spread.

These precautionary behavioral changes during January and February 2020 showed a more relaxed restriction than formal social distancing rules or other non-pharmaceutical interventions (NPIs) implemented later (e.g., the first group gathering ban was effective from March 29, 2020). By comparing the effective reproduction number (R_t) in influenza along with these differences in the precautionary behaviors, the corresponding effects can be quantified, which provide important insights to understand whether an influenza outbreak can be controlled using less intensive social distancing restrictions without huge socioeconomic impacts. Furthermore, whether individuals practice precautionary measures or choose to be vaccinated largely depends on their risk perception, relating to a complex decision-making process (18, 19). Knowing the impacts of these behavioral changes help to forecast the possible epidemic situations after reopening.

Timely public health decision-making often needs to be made during an outbreak. However, the methods of estimating parameters, such as R_t, of traditional susceptible-infectious-recovered (SIR) equations under Bayesian framework (e.g., Markov chain Monte Carlo (20), particle filtering (21, 22), etc.) are usually time-consuming. Alternatively, the time-series susceptible-infectious-recovered (TS-SIR) model provides a computationally inexpensive way to model the transmission dynamics as the parameters can be estimated through a generalized linear model (GLM) (23–25). Compared with frequentist approaches, Bayesian approaches to modeling and inference of infectious disease dynamics have the advantage that latent parameters (e.g., actual numbers of uninfected (susceptible) and infected individuals) and their uncertainties can be seamlessly accounted for (26). To further reduce the computational load from traditional methods for Bayesian inference, some approximation methods such as integrated nested Laplace approximations (INLA) approach can be applied (27).

The aim of our study was to identify the relationship between precautionary behaviors (e.g., wearing face masks and avoiding crowded places) and the reduction in influenza transmissibility. We adopted a TS-SIR model to estimate the R_t in influenza seasons by using a Bayesian approach with INLA. TS-SIR was transformed to a GLM with Poisson regression. After considering the effect of underreporting and separating the effect from herd immunity, we were able i) to quantify the effects of wearing face masks and avoiding crowded places throughout an outbreak and ii) to identify the required percentage of people adopting such precautionary behaviors that could suppress the outbreak.

2. Materials and methods

2.1. Data collection

The weekly reported influenza cases in Hong Kong from April 12, 2015 to March 22, 2020 were obtained from the Centre for Health Protection (CHP) (28). Only outbreaks during regular winter seasons were collected for our study (Figure 1A).

2.2. Modeling

On January 14, 2020, WHO made an announcement of COVID-19 outbreak, and shortly afterward, China declared a first-level public health emergency response (15, 17). Hence,

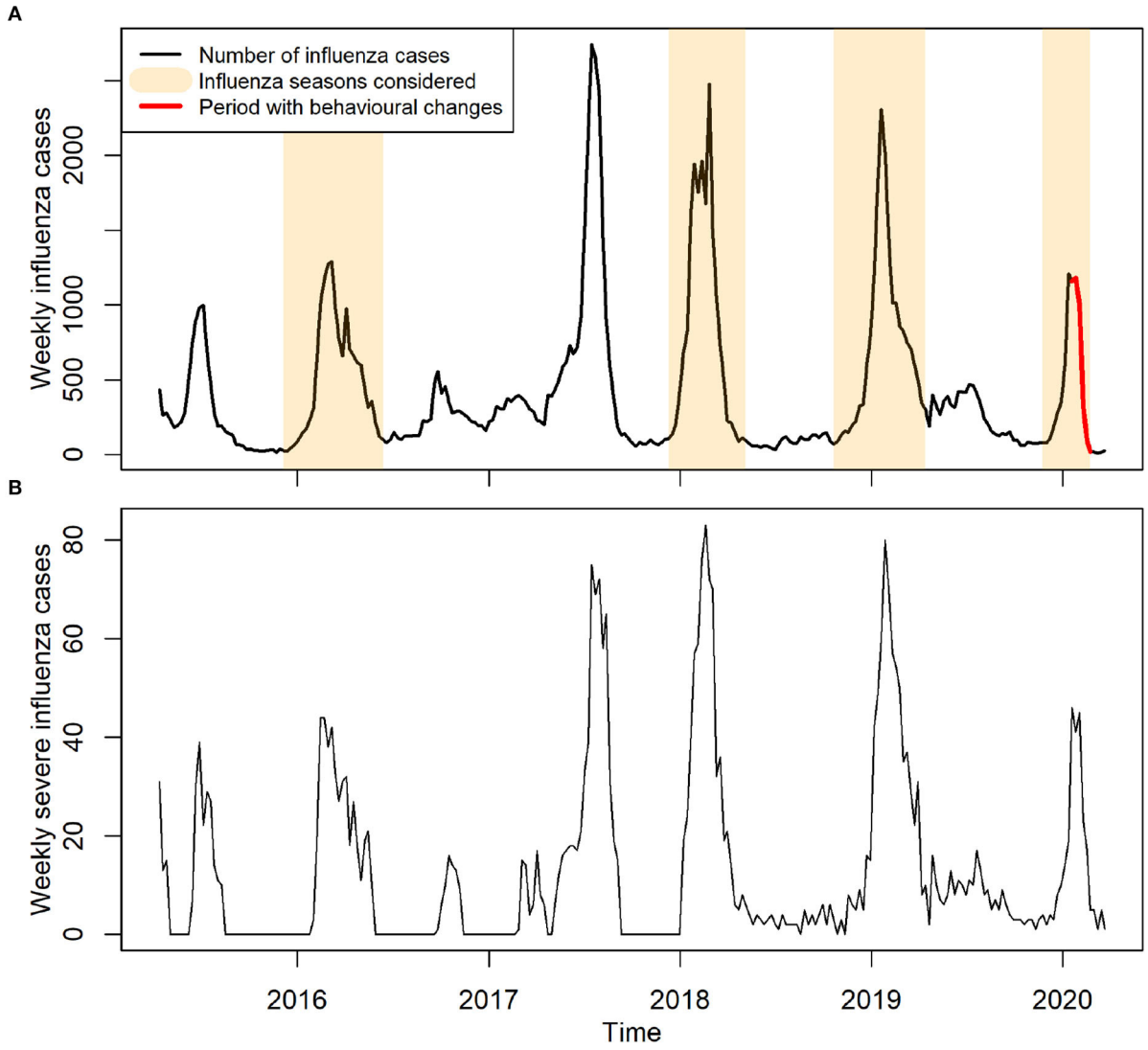


FIGURE 1
Reported cases of influenza in Hong Kong between years 2015 and early 2020. **(A)** Weekly reported influenza cases. Influenza seasons considered for model training are highlighted in light brown color, indicating an ordinary influenza period. Red color indicates the period with precautionary behavioral changes. **(B)** Weekly reported severe influenza cases.

within the study period of 2019/20 seasonal influenza outbreak (Figure 1A), we defined the period from the start of the outbreak (November 24, 2019) up to January 12, 2020 before the majority of Hong Kong residents knew the existence of COVID-19 as an ordinary influenza period. This ordinary period also included the winter influenza seasons during the years 2015–2016, 2017–2018, and 2018–2019. We included these previous seasons to increase the statistical power and obtain robust estimates of the baseline reporting rate. We did not consider the year 2016–2017 in our training set as there was no obvious winter seasonal peak.

Three phases with different transmission patterns were observed in the 2019–2020 winter influenza season (i.e., the growing, plateau, and decline periods). These phases were correlated to the stages of COVID-19 spread in Hong Kong. The **ordinary phase** (Phase 1) indicated the period before COVID-19. In addition to the ordinary phase (corresponding to the growing period of the outbreak), we further split the influenza season after January 12, 2020 into two other phases: the **awareness phase** (Phase 2), from January 12 to February 2, when people received the announcement given by WHO, corresponding to the plateau period; and the **spreading phase** (Phase 3), from February 2 to February 23, during which local community transmission occurred in Hong Kong, corresponding to the decline period.

Time-Series Susceptible-Infectious-Recovered (TS-SIR) model

Effective reproduction number R_t at time t was calculated as follows:

$$R_t = R_0 \times C_j \times (\frac{S_t}{N}), \tag{1}$$

where R_0 is the basic reproduction number and C_j is the contact ratio, a ratio of the contact rate during phase j compared to the pre-pandemic period (Phase 1; $j = 1$). The baseline contact ratio for the pre-pandemic period was fixed at $C_1 = 1$. S_t is the susceptible population at time t and N is the population in Hong Kong. Since official population data were only reported in 2016 and 2021 from the Census and Statistics Department of Hong Kong (29), we assumed the population growth to be linear between 2016 and 2021, that is, the population was 7,336,585 in 2016, 7,365,931 in 2018, 7,380,605 in 2019, and 7,395,278 in 2020. We defined the effects of behavioral changes on transmissibility in phase j (Φ_j) as the percentage of reduction in contact ratio, in which the effect of herd immunity (i.e., the effect contributed by the reduction in susceptible population over time) is removed:

$$\Phi_j = (1 - C_j) \times 100\%. \tag{2}$$

While many of the epidemiological models used for influenza modeling are conventional compartmental models (i.e., SIR model), an alternative, though related, model is the TS-SIR model (24), which transforms the conventional model to a GLM, a classic regression approach. In this study, we adopted a TS-SIR model with reference to Imai et al. (24) to capture the transmission dynamics for influenza and we considered different reporting rates at different periods of time due to public's risk perception amid COVID-19:

$$\begin{aligned} \frac{Y_t}{\rho_j} &= R_{t-1}^{T_c} \times \frac{Y_{t-1}}{\rho_j} \\ &= (R_0 \times C_j \times \frac{S_{t-1}}{N})^{T_c} \times \frac{Y_{t-1}}{\rho_j} \\ &= (R_0 \times C_j \times \frac{N - \sum_{i=0}^{t-1} Y_i/\rho_j}{N})^{T_c} \times \frac{Y_{t-1}}{\rho_j}, \end{aligned} \tag{3}$$

where Y_t is the reported incidence at time t and S_{t-1} is the susceptible population at time $t - 1$; we considered the susceptible population equal to total population minus the cumulative incidence within an influenza season, i.e., $S_{t-1} = N - \sum_{i=0}^{t-1} Y_i/\rho_j$, ρ_j is the reporting rate at Phase j. Because weekly influenza data are published by CHP in Hong Kong, the unit of t is week (and $t = 1, 2, ...$). To calculate the number of infected cases generated from a single infected case after a unit of time, a time scale T_c relative to the generation time of 3.5 days is introduced (30), which is calculated as 7/3.5 = 2.

Equation (3) can be transformed to a GLM with Poisson distribution (see Supplementary material for details), such that $Y_t/\rho_j \sim \text{Poisson}(\mu_t)$, where μ_t denotes the expected value of the weekly influenza cases. We had $\log(\mu_t) = \log(\frac{Y_{t-1}}{\rho_j}) + T_c \log(R_0 \times C_j) - \frac{T_c \times \sum_{i=0}^{t-1} Y_i/\rho_j}{N}$.

To estimate parameters, we first obtained R_0 and ρ_1 during the ordinary period (which includes the winter influenza seasons during 2015–16, 2017–18, 2018–19, and 2019–20 up to January 12, 2020). Then, we obtained ρ_2, C_2, ρ_3, and C_3, subsequently, by modeling the situations in Phase 2 and Phase 3. The detailed procedures for model fitting can be found in Supplementary material.

We estimated the effects of wearing face masks and avoiding crowded places on the contact ratio defined in our model. The percentage of reduction in contact ratio was used to represent the percent reduction in R_t while excluding the impact from the herd immunity (see Equation 2). We assumed that these two effects are independent and additive; thus, we have

$$\Phi_j = \phi_{sd}(x_{j,sd} - x_{1,sd}) + \phi_m(x_{j,m} - x_{1,m}), \tag{4}$$

where Φ_j is the overall percent reduction of contact ratio in Phase 2 and Phase 3 ($j = 2, 3$), which was previously estimated from Equations 1 and 2; $x_{j,sd}$ and $x_{j,m}$ are the percentages of population avoiding crowded places and wearing face masks in Phase j, respectively. $x_{1,sd}$ and $x_{1,m}$ are their baseline percentages (the estimates of the baseline percentages come from our survey, see Results Section for details). ϕ_{sd} and ϕ_m are parameters indicating the effectiveness of individual behaviors. The product of ϕ and x was referred as the effect on contact ratio in total population. To account for the uncertainty in the extent of avoiding crowded places and wearing face masks at different periods of time, we assumed that the number of survey participants avoiding crowded places or wearing face masks followed a binomial distribution (the number of trials is equal to the population in Hong Kong in 2020, with different probabilities in the different phases according to the mean percentages in the surveys. We adopted a bootstrap approach to capture the uncertainty in model parameters (see Supplementary material for details). The code for the abovementioned models can be found at https://github.com/hy39/ts-sir-flu.

3. Results

To quantify the effects of behavioral changes (in response to the initial spread of COVID-19) on influenza transmissibility, we classified the 2019–20 winter influenza season into three different phases based on the pattern of influenza activity as mentioned in the Section 2 (Figure 2): Phases 1, 2, and 3 show the growth, plateau, and decline phases of the dynamics, respectively. Comparing influenza activity in year 2019–2020 with the previous seasons, the growth became apparently limited after Phase 1 and then reduced significantly without

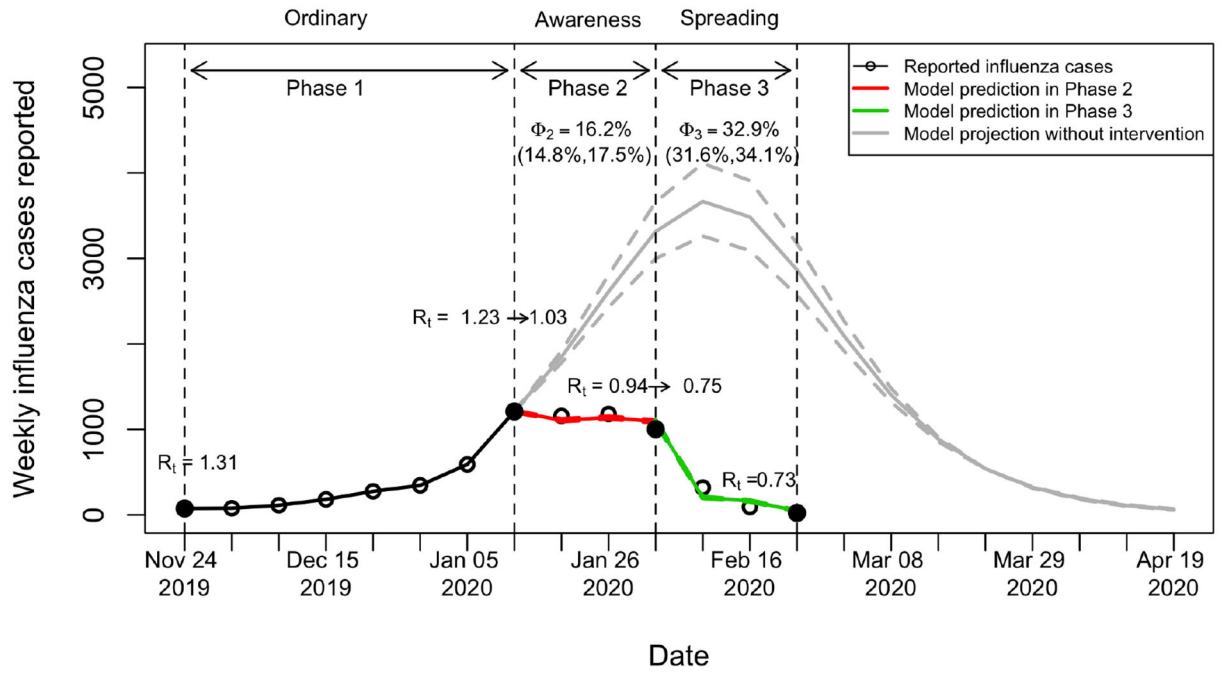


FIGURE 2
Model prediction of influenza cases in 2019–20 winter influenza season in Hong Kong. R_t at the boundaries (the second and third dashed lines) between phases were estimated from the model taking account of changes in both behavior and herd immunity. The prediction in Phase 2 (Awareness: the period immediately after people were aware of the existence of COVID-19) is shown in red. The prediction in Phase 3 (Spreading: the period immediately after local COVID-19 cases began to spread) is shown in green. The prediction intervals in Phase 2 and 3 are shown with narrow intervals. Φ denotes the percentage of reduction in R_t at Phase 2 and 3 compared to Phase 1, resulting from the changes in behavior only (i.e., excluding the effects from herd immunity).

showing a typical curvature of epidemic peak. Presumably, this unusual pattern was due to the human behavioral changes associated with people's risk perception on certain critical public health events (i.e., COVID-19 spreading) (Figure 3B). Hence, we correlated these three phases to different epidemic statuses of COVID-19, namely ordinary, awareness, and spreading phases (see Section 2 and Figure 2).

To estimate the effects of behavioral changes on the transmissibility, we adopted a TS-SIR model by taking account of the herd immunity changes. Our model captured the dynamics across the three phases well. The number of influenza cases stopped growing after people avoided crowded places and wore face masks. Compared with the projection of cases without the effects of behavioral changes (i.e., under ordinary transmission dynamics), the number of cases began to decline at least 4 weeks earlier and the total number of reported cases until February 23 was reduced by 78.8% (Figure 2).

Initial R_t, also called the basic reproduction number R_0, was estimated to be 1.37 (95%CI, 1.35–1.4). In 2020 winter influenza season, the R_t reduced slightly from 1.31 to 1.23 during Phase 1 (Figures 2, 3A), which was mainly caused by the increase in herd immunity after the infected cases were recovered. After the risk of COVID-19 transmission has been noticed, the R_t on January 12 reduced from 1.23 (end of Phase 1) to 1.03 (start of Phase 2), with an effect of behavioral changes Φ_2 (i.e., the percentage of reduction in R_t after excluding the effects of herd immunity; see Section 2) being 16.2% (95%CI, 14.8–17.5%) in this phase. The R_t on February 2 further reduced from 0.94 (end of Phase 2) to 0.75 (start of Phase 3), with an overall effect of behavioral changes Φ_3 being 32.9% (95%CI, 31.6–34.1%) in this phase. In Phase 3, the R_t slightly reduced to 0.73 at the end. The prediction intervals in Phase 2 and Phase 3 are narrow because the uncertainty of adjusted reporting rates at the corresponding phases was small (see Supplementary material and subsequent paragraphs for details). The results showed that the reductions in transmissibility were primarily due to the behavioral changes against COVID-19 and only partially due to the increase in herd immunity (Figure 3A).

We next quantified the effects of different behaviors. The survey during the baseline showed that 37.9% of people would avoid crowded places and 45.5% would wear face masks for preventing influenza infection (Figure 4). The behavioral changes in the subsequent phases were revealed by the study of Cowling et al. (16) (Figure 4), in which two surveys were conducted immediately after Phase 2 and 3 began (Figure 3A). The percentage of people avoiding crowded places increased from 60% to 90%, while the percentage for wearing face masks increased from 75 to 98%. At that time, social distancing rules,

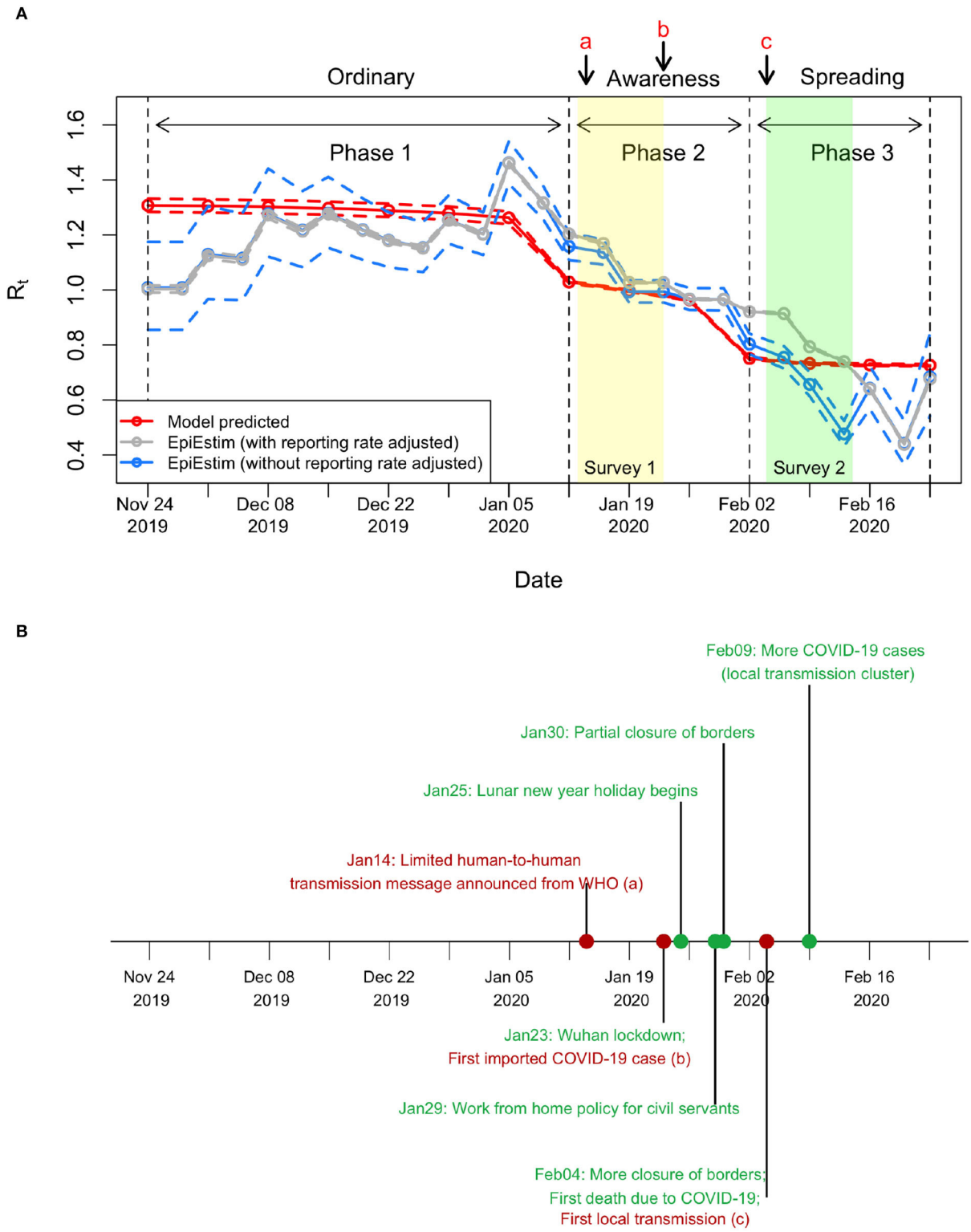


FIGURE 3
Changes in R_t before COVID-19 outbreak. **(A)** Estimated R_t in 2019/20 winter influenza season. The mean R_t was estimated as 1.29 (95%CI, 1.27–1.32) in Phase 1, reduced to 1.00 (95%CI, 0.99–1.00) in Phase 2, and further reduced to 0.73 (95%CI, 0.73–0.74) in Phase 3. In applying EpiEstim package, the number of influenza cases per week was converted into cases per 3.5 days by linear interpolation on the cumulative influenza cases, which equals to roughly one generation time of influenza (30). For other settings in applying the EpiEstim package, the serial interval was a gamma distribution with the mean equal to 3.5 days and the standard deviation equal to 1 day, and the window size was 2 weeks. **(B)** Timeline about COVID-19. Important events are shown in red, while other events are shown in green.

TABLE 1 Impacts of precautionary behaviors in influenza control and transmissibility from our model.

Impact	Value (95%CI)
(%) Effectiveness of avoiding crowded places	44.1 (43.5, 44.7)
(%) Effectiveness of wearing face masks	17.4 (16.3, 18.3)
Estimated R_0 when nobody wears face masks or avoids crowded places	1.71 (1.707, 1.715)
Percentage of people required to adopt precautionary behaviors in order to lower R_0 to below 1	84.3% (84.1%, 84.5%)

such as gathering ban, have not been implemented by the Hong Kong government yet. Both avoiding crowded places and wearing face masks were precautionary behaviors triggered by individual's risk perception.

We estimated the effects of these two behavioral changes after separating the effect of herd immunity (see Section 2). The results showed that wearing face masks was associated with an 17.4% reduction in R_t (the coefficient is 0.174, 95%CI, 0.163–0.183) (Table 1). The effect of avoiding crowded places was 44.1% (the coefficient is 0.441, 95%CI, 0.435 to 0.447). When nobody wears face masks or avoids crowded places, the initial R_t was 1.71 (95%CI, 1.707–1.715), which was higher than our previous estimate of 1.37. This is because a fraction of people have adopted preventive measures for influenza (Figure 4). To reduce R_0 to below one, more than 84.3% (95%CI, 84.1–84.5%) of people have to wear face masks and avoid crowded places.

Note that we addressed the concerns of underreporting due to COVID-19 by adjusting the reporting rates in different phases (see Supplementary material for details) with the ratio of the severe influenza cases to total reported influenza cases (Table 2 and Figure 1B). The ordinary reporting rate was estimated as 0.0065 (95%CI, 0.0064–0.0067), and the adjusted reporting rate dropped to 0.0057 (95%CI, 0.0056–0.0059) in Phase 2 and further to 0.0022 (95%CI, 0.0021–0.0022) in Phase 3. There was a reduction in reporting rate (0.0008 and 0.0043, respectively, for Phase 2 and Phase 3, compared with Phase 1) across the three phases, which conforms to the expectation that fewer patients with influenza visit hospitals or clinics under the risk of COVID-19.

Our results were compared with the R_t estimated using data on the number of observed new cases with a statistical method based on renewal function (EpiEstim package) (31) (Figure 3A). The comparison showed that the R_t estimations from both methods were consistent. However, the R_t from the EpiEstim package showed larger variations within each phase than our predictions. Before the spread of COVID-19 (i.e., Phase 1), the R_t from EpiEstim are similar with and without reporting rate adjusted (gray line and blue line, respectively, in Figure 3A). However, without adjusting reporting rate, the R_t from EpiEstim was lower than the R_t estimates with the reporting rate adjusted.

4. Discussion

The importance of wearing face masks on stopping COVID-19 spread through droplet or aerosol transmission has been addressed by the WHO (32). Although the effects of wearing face masks on preventing common respiratory virus infection, such as influenza or SARS-CoV-2, have been intensely discussed using empirical evidence from laboratory (33–36) or simulation studies (37), the evidence from the population study is little (38). We quantified the effects of wearing face masks and avoiding crowded places on seasonal influenza transmissibility during early COVID-19 spread period when human behavior changed. The results demonstrated that precautionary behavioral changes may have had a large impact on influenza transmission, even before strict social distancing rules were implemented. This gives important recommendation on the prevention of future influenza using NPIs.

Possibly because of the risk perception related to previous experience of SARS epidemic in 2003, the adoption rate of face masks in Hong Kong was high even before COVID-19 began to spread in the community [see Figure 4 from our data and the data revealed by the surveys from Cowling et al. (16) and Kwok et al. (39)]. Even though these spontaneous behavioral changes were less restricted than formal social distancing rules, our estimation, taking account of the underreporting of influenza cases, showed that R_t reduced from 1.31 to 0.73 (Figure 2). High risk perception may also affect the decision-making in vaccination. The complex relationships between behavioral changes and transmission dynamics can be modeled through the evolutionary game theory (18, 19), which is important in predicting and preparing for future outbreaks after reopening.

A reduction in the incidence of influenza has also been reported in mainland China during the early COVID-19 spread (40), which further supports the finding that the interventions implemented against COVID-19 significantly reduced influenza incidence. The interventions appeared to have different degrees of impact on influenza incidence than in Hong Kong, which was likely because of the start time of the influenza season. In mainland China, the influenza season began in November 2019 (40) but the COVID-19 interventions (first-level responses) were implemented in late January 2020, when the epidemic peak has already been reached. However, Hong Kong had an influenza season at a later time and the COVID-19-induced interventions or precautionary behaviors were adopted before the peak (Figure 2). Hence, the incidence of influenza incidence was less in Hong Kong.

Based on the data from Hong Kong, our results showed that wearing face masks could reduce seasonal influenza transmission by as much as 17.4% in the population, which is nearly half of the effect of avoiding crowded places (44.1%). Within our model, if more than 85% of the people had avoided crowded places and wore face masks, R_t could have reduced to

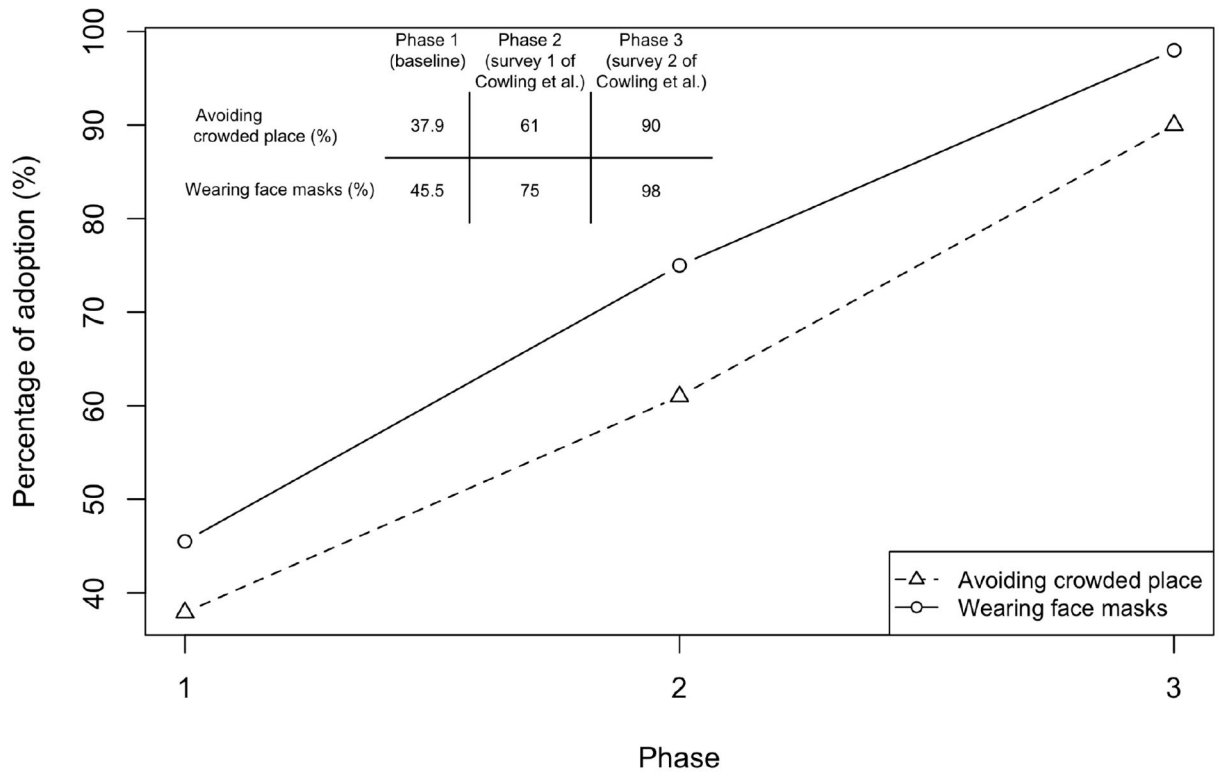


FIGURE 4
Changes in behavior in response to COVID-19. Results of Phase 1 were taken from a survey conducted during January 5–February 15, 2020 for assessing the baseline behavior taken to prevent influenza infection. Sample size is 66, with 45 females. Respondents were aged between 19 and 64. Three respondents have taken flu vaccination. Note that the survey questions only ask the possible measures in order to prevent influenza but not COVID-19. Hence, the results indicate the baseline behavior before COVID-19 emerged (Avoiding crowded places: 37.9%, 95%CI, 26.2–49.6%; Wearing face masks: 45.5%, 95%CI, 33.5–57.5%). Results of Phases 2 (Cowling Survey 1) and 3 (Cowling Survey 2) were taken from a previous study conducted by Cowling et al. (16)) before NPIs were implemented in Hong Kong. Cowling Survey 1 (January 20-23) was conducted immediately after the WHO made an announcement on January 14, 2020 and when China declared a first-level public health emergency response (January 20, 2020) (15, 17) (sample size is 1008; Avoiding crowded places: 61%, 95%CI, 57.2–65.4%; Wearing face masks: 75%, 95%CI, 70.4–78.6%). Cowling Survey 2 (February 11-14) was conducted after the first community transmission event was confirmed in Hong Kong (sample size is 1000; Avoiding crowded places: 90%, 95%CI, 86.2–94.2%; Wearing face masks: 98%, 95%CI, 93.5–100%).

TABLE 2 Ratios of severe cases to the total reported cases in several winter influenza seasons.

Year	Ratio (%)
2016	3.1
2018	3.5
2019	3.8
2020 - Phase 2	4.0
2020 - Phase 3	10.0

The ratios were applied to adjust the reporting rates in Phase 2 and Phase 3 of 2019/2020 winter influenza season.

below one. This evidence suggests that, without strong policy restrictions in social distancing (i.e., four-person gathering ban in public places or even a lockdown), the incidence of influenza can still be greatly reduced by simple behavioral changes. This highlights the need of future research on whether mandatory mask wearing policy in certain public places only (e.g., public transport or other crowded places) can significantly reduce influenza infection.

A TS-SIR model provides a convenient way to estimate epidemiological parameters using the classical GLM approach without losing the nonlinear effects in the conventional SIR model. To estimate the reproduction number accurately, our model took account of the change in reporting rate due to the outbreaks of COVID-19, with reference to the fact that people were reluctant to go to the clinic (81% (16) and 76% (39) of the respondents). Using the TS-SIR model, we were able to separate the changes in R_t due to both behavioral changes and the increase in herd immunity. This allowed us to quantify the changes in R_t caused by the behavioral changes using a classical statistical approach.

Nevertheless, there are some limitations in our study. In the proposed model, we assumed the population was random mixing without considering the effect of heterogeneous mixing (24). We mainly assumed that avoiding crowded places and wearing face masks were the major behaviors

that could reduce the number of transmissions, because these transmissions mainly occur through the droplets released when an infected person sneezes, coughs, or speaks. In addition, in the effectiveness-behavior analysis (Equation 4), we checked whether confounding factors occurred. We found that the probability of wearing face masks is not conditional on avoiding crowded places, enabling us to use a simple additive approach to assess the individual effects. A larger sample indeed can increase the statistical power of the effectiveness-behavior analysis (i.e., the credible intervals of the resulted effects). However, due to time constraints, the sample size in our survey on Phase 1 behavior was limited. Furthermore, the surveyed behavioral changes were simply interpreted in the percentage of the population. Future studies should be conducted to quantify the duration of wearing face masks.

Preparing for the co-circulation of influenza and COVID-19 is critically important (14). While many countries lifted the requirements of social distancing and wearing face masks as COVID-19 vaccination rolled out and the omicron (B.1.1.529) outbreaks passed, these relaxations likely led to the rise of seasonal influenza infections. However, COVID-19 continues to spread with more than 1 million new cases per day globally (March 2022) (41). Some biotechnology companies have been developing a COVID-19-Influenza combination vaccine to provide protections from both illnesses at the same time (42, 43). To reduce the influenza incidence from a non-pharmaceutical perspective, we recommend that the idea of wearing face masks in certain public places and/or simple social distancing (i.e., avoiding crowded places) should be promoted. The effectiveness of such precautionary behaviors on seasonal influenza based on our study can also potentially give us hints for the recommendations of behavioral shift in dealing with future pandemics (44).

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author. The code for the abovementioned models can be found at https://github.com/hy39/ts-sir-flu.

Author contributions

H-YY and C-PL contributed to conception and design of the study and wrote the first draft of the manuscript. C-PL performed the analysis. ID, K-LT, MX, and M-HL wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This research was funded by the Health and Medical Research Fund (COVID190215), the City University of Hong Kong (7200573 and 9610416), the Wellcome Trust and the Royal Society (213494/Z/18/Z), and Hong Kong Innovation and Technology Commission (InnoHK Project CIMDA).

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.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.992697/full#supplementary-material

References

1. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*. (2020) 368:860–8. doi: 10.1126/science.abb5793

2. Charumilind S, Craven M, Lamb J, Sabow A, Wilson M. *When Will the COVID-19 Pandemic End*? (2020). Available online at: https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/when-will-the-covid-19-pandemic-end#.

3. Scudellari M. How the pandemic might play out in 2021 and beyond. *Nature*. (2020) 584:22–5. doi: 10.1038/d41586-020-02278-5

4. Freedman DH. *Winter Will Make the Pandemic Worse. Heres What You Need to Know*. MIT Technology Review (2020). Available online at: https://www.technologyreview.com/2020/10/08/1009650/winter-will-make-the-pandemic-worse/.

5. Ting V. *Hong Kong Facing 1,000 Winter Deaths From Coronavirus, Expert Warns, as City Records Seven New Cases of COVID-19*. South China Morning Post (2020). Available online at: https://www.scmp.com/news/hong-kong/health-environment/article/3105022/hong-kong-facing-1000-winter-deaths-coronavirus/.

6. de Celles MD, Casalegno JS, Lina B, Opatowski L. Influenza may facilitate the spread of SARS-CoV-2. *medRxiv*. (2020). doi: 10.1101/2020.09.07.20189779

7. Yue H, Zhang M, Xing L, Wang K, Rao X, Liu H, et al. The epidemiology and clinical characteristics of co-infection of SARS-CoV-2 and influenza viruses in patients during COVID-19 outbreak. *J Med Virol*. (2020) 92:2870–2873. doi: 10.1002/jmv.26163

8. Hagen A. *COVID-19 and the Flu*. American Society for Microbiology (2020). Available online at: https://asm.org/Articles/2020/July/COVID-19-and-the-Flu.

9. Zhang L, Zhang Y. Influenza viral infection is a high-risk factor for developing coronavirus disease 2019 (COVID-19). *Preprints*. (2020) 2020:2020030307. doi: 10.20944/preprints202003.0307.v1

10. Lin II RG, Money L. *California Sees Record-Breaking COVID-19 Deaths, A Lagging Indicator of Winter Surge*. Los Angeles Times (2021). Available online at: https://www.latimes.com/california/story/2021-01-22/california-sees-record-breaking-covid-19-deaths-a-lagging-indicator-of-winter-surge.

11. Coren MJ. *The Coronavirus in Winter May be Worse Than Scientists Thought*. Quartz (2021). Available online at: https://qz.com/1961313/the-coronavirus-in-winter-may-be-worse-than-scientists-thought/.

12. Roache M. *COVID-19 Cases Are Rising in Europe, and the Vaccine Rollout Is Lagging. How Did It All Go So Badly Wrong*? Time (2021). Available online at: https://time.com/5945220/europe-covid-19-surge-2021/.

13. World Health Organization. *Preparing for the Next Human Influenza Pandemic: Celebrating 10 Years of the Pandemic Influenza Preparedness Framework*. (2021). Available online at: https://www.who.int/news-room/feature-stories/detail/preventing-the-next-human-influenza-pandemic-celebrating-10-years-of-the-pandemic-influenza-preparedness-framework/.

14. World Health Organization. *GISRS for the Upcoming Influenza Seasons During the COVID-19 Pandemic*. (2020). Available online at: https://www.who.int/influenza/surveillance_monitoring/gisrs_covid19_upcoming_flu_season/en/.

15. Nebehay S. *WHO Says new China Coronavirus Could Spread, Warns Hospitals Worldwide*. Reuters (2020). Available online at: https://www.reuters.com/article/us-china-health-pneumonia-who-idUSKBN1ZD16J.

16. Cowling BJ, Ali ST, Ng TW, Tsang TK, Li JC, Fong MW, et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. *Lancet Public Health*. (2020) 5: e279–88. doi: 10.1016/S2468-2667(20)30090-6

17. Yu X, Li N. Understanding the beginning of a pandemic: Chinas response to the emergence of COVID-19. *J Infect Public Health*. (2020) 14:347–52. doi: 10.1016/j.jiph.2020.12.024

18. Tanimoto J. *Sociophysics Approach to Epidemics. Vol. 23*. Singapore: Springer (2021).

19. Tanimoto J. *Evolutionary Game Theory With Sociophysics: Analysis of Traffic Flow and Epidemics. Vol. 17*. Singapore: Springer (2018).

20. Thompson R, Stockwin J, van Gaalen RD, Polonsky J, Kamvar Z, Demarsh P, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*. (2019) 29:100356. doi: 10.1016/j.epidem.2019.100356

21. Liang J, Yuan HY, Wu L, Pfeiffer DU. Estimating effects of intervention measures on COVID-19 outbreak in Wuhan taking account of improving diagnostic capabilities using a modelling approach. *BMC Infect Dis*. (2021) 21:1–10. doi: 10.1186/s12879-021-06115-6

22. Endo A, Van Leeuwen E, Baguelin M. Introduction to particle Markov-chain Monte Carlo for disease dynamics modellers. *Epidemics*. (2019) 29:100363. doi: 10.1016/j.epidem.2019.100363

23. Becker AD, Grenfell BT. tsiR: an R package for time-series Susceptible-Infected-Recovered models of epidemics. *PloS One*. (2017) 12:e0185528. doi: 10.1371/journal.pone.0185528

24. Imai C, Armstrong B, Chalabi Z, Mangtani P, Hashizume M. Time series regression model for infectious disease and weather. *Environ Res*. (2015) 142:319–327. doi: 10.1016/j.envres.2015.06.040

25. Koelle K, Pascual M. Disentangling extrinsic from intrinsic factors in disease dynamics: a nonlinear time series approach with an application to cholera. *Am Natural*. (2004) 163:901–13. doi: 10.1086/420798

26. Chatzilena A, van Leeuwen E, Ratmann O, Baguelin M, Demiris N. Contemporary statistical inference for infectious disease models using Stan. *Epidemics*. (2019) 29:100367. doi: 10.1016/j.epidem.2019.100367

27. 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) 71:319–92. doi: 10.1111/j.1467-9868.2008.00700.x

28. Centre for Health Protection, *Department of Health, Hong Kong SAR*. Flu express (2020). Available online at: https://www.chp.gov.hk/en/resources/29/304.html.

29. Census and Statistics Department. *2021 Population Census. Census and Statistics Department*. (2021). Available online at: https://www.censtatd.gov.hk/en/scode600.html.

30. Cowling BJ, Fang VJ, Riley S, Peiris JM, Leung GM. Estimation of the serial interval of influenza. *Epidemiology (Cambridge, Mass)*. (2009) 20:344. doi: 10.1097/EDE.0b013e31819d1092

31. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am J Epidemiol*. (2013) 178:1505–12. doi: 10.1093/aje/kwt133

32. World Health Organization. *Transmission of SARS-CoV-2: Implications for Infection Prevention Precautions: Scientific Brief, 09 July (2020)*. World Health Organization (2020). Available online at: https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions

33. Leung NH, Chu DK, Shiu EY, Chan KH, McDevitt JJ, Hau BJ, et al. Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat Med*. (2020) 26:676–80. doi: 10.1038/s41591-020-0843-2

34. Kelland K. *Masks do Reduce Spread of Flu and Some Coronaviruses, Study Finds*. Reuters (2020). Available online at: https://www.reuters.com/article/uk-health-coronavirus-masks-science/masks-do-reduce-spread-of-flu-and-some-coronaviruses-study-finds-idUKKBN21L2I1.

35. Chan JFW, Yuan S, Zhang AJ, Poon VKM, Chan CCS, Lee ACY, et al. Surgical mask partition reduces the risk of non-contact transmission in a golden Syrian hamster model for coronavirus disease 2019 (COVID-19). *Clin Infect Dis*. (2020). 71:2139–49. doi: 10.1093/cid/ciaa644

36. Prather KA, Wang CC, Schooley RT. Reducing transmission of SARS-CoV-2. *Science*. (2020) 368:1422–4. doi: 10.1126/science.abc6197

37. Stutt RO, Retkute R, Bradley M, Gilligan CA, Colvin J. A modelling framework to assess the likely effectiveness of facemasks in combination with 'lock-down in managing the COVID-19 pandemic. *Proc R Soc A*. (2020) 476:20200376. doi: 10.1098/rspa.2020.0376

38. Rader B, White LF, Burns MR, Chen J, Brilliant J, Cohen J, et al. Mask-wearing and control of SARS-CoV-2 transmission in the USA: a cross-sectional study. *Lancet Digit Health*. (2021) 3:e148-e157. doi: 10.1016/S2589-7500(20)30293-4

39. Kwok KO, Li KK, Chan HH, Yi YY, Tang A, Wei WI, et al. Community responses during early phase of COVID-19 epidemic, Hong Kong. *Emerg Infect Dis*. (2020) 26:10–3201. doi: 10.3201/eid2607.200500

40. Kong X, Liu F, Wang H, Yang R, Chen D, Wang X, et al. Prevention and control measures significantly curbed the SARS-CoV-2 and influenza epidemics in China. *J Virus Eradicat*. (2021) 7:100040. doi: 10.1016/j.jve.2021.100040

41. World Health Organization. *WHO Coronavirus Disease (COVID-19) Dashboard*. (2022). Available online at: https://covid19.who.int/

42. Novavax. *Initial Results From Novavax' COVID-19-Influenza Vaccine Trial Are First to Show Feasibility of Combination Vaccine*. Novavax (2022). Available online at: https://ir.novavax.com/2022-04-20-Initial-Results-from-Novavax-COVID-19-Influenza-Vaccine-Trial-are-First-to-Show-Feasibility-of-Combination-Vaccine

43. Hopkins J. *Moderna Eyes Dual COVID-19, Flu Vaccine*. The Wall Street Journal (2021). Available online at: https://www.wsj.com/livecoverage/covid-2021-03-23/card/qVE01KdPTaZWp3oBX9yb

44. Griffin M, Sohrabi C, Alsafi Z, Nicola M, Kerwan A, Mathew G, et al. Preparing for COVID-19 exit strategies. *Ann Med Surg*. (2020) 61:88–92. doi: 10.1016/j.amsu.2020.12.012

frontiers | Frontiers in Public Health

TYPE Brief Research Report
PUBLISHED 01 December 2022
DOI 10.3389/fpubh.2022.1002440

OPEN ACCESS

EDITED BY
Theodore Gyle Lewis,
Naval Postgraduate School,
United States

REVIEWED BY
Gillie Gabay,
Achva Academic College, Israel
Pierpaolo Ferrante,
National Institute for Insurance against Accidents at Work (INAIL), Italy

*CORRESPONDENCE
Royi Barnea
royib@assuta.co.il

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Public Health Policy, a section of the journal Frontiers in Public Health

RECEIVED 25 July 2022
ACCEPTED 18 November 2022
PUBLISHED 01 December 2022

CITATION
Niv-Yagoda A, Barnea R and Rubinshtein Zilberman E (2022) The role of models as a decision-making support tool rather than a guiding light in managing the COVID-19 pandemic. *Front. Public Health* 10:1002440. doi: 10.3389/fpubh.2022.1002440

The role of models as a decision-making support tool rather than a guiding light in managing the COVID-19 pandemic

Adi Niv-Yagoda[1,2†], Royi Barnea[1,3*†] and Efrat Rubinshtein Zilberman[4]

[1]School of Health Systems Management, Netanya Academic College, Netanya, Israel, [2]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel, [3]Assuta Health Services Research Institute, Assuta Medical Centers, Tel-Aviv, Israel, [4]Hillel Yaffe Medical Center, Hadera, Israel

Reference scenarios based on mathematical models are used by public health experts to study infectious diseases. To gain insight into modeling assumptions, we analyzed the three major models that served as the basis for policy making in Israel during the COVID-19 pandemic and compared them to independently collected data. The number of confirmed patients, the number of patients in critical condition and the number of COVID-19 deaths predicted by the models were compared to actual data collected and published in the Israeli Ministry of Health's dashboard. Our analysis showed that the models succeeded in predicting the number of COVID-19 cases but failed to deliver an appropriate prediction of the number of critically ill and deceased persons. Inherent uncertainty and a multiplicity of assumptions that were not based on reliable information have led to significant variability among models, and between the models and real-world data. Although models improve policy leaders' ability to act rationally despite great uncertainty, there is an inherent difficulty in relying on mathematical models as reliable tools for predicting and formulating a strategy for dealing with the spread of an unknown disease.

KEYWORDS

COVID-19, models, health policy, evidence based decision-making, public health

Introduction

By its very nature, health is a statistical science replete with uncertainty, which is particularly high in certain situations such as pandemics. Hence over the years, and in various situations, decision makers in different countries have sought the help of reference scenarios and models to predict disease spread as a tool for effective ways to prevent it and for formulating tailored health policies (1–3). This was also the case when the magnitude of local infection with COVID-19 in the Chinese province of Wuhan came to light and morbidity spread to other countries until the WHO declared it a pandemic (4).

With the increasing reports about the spread of the disease, and like many countries worldwide, Israel adopted the recommendations of the WHO and began to take behavioral-social preventive measures in order to slow down and reduce the rate of COVID-19 spread. The measures described below were incorporated into Israel's health policy, some of them were unique to Israel.

Preventive measures taken before the first COVID-19 wave

In January 2020, health institutions were instructed to be vigilant regarding individuals who returned from China and showed symptoms of illness (fever, cough, etc.), and to increase awareness among medical teams, and the importance of protecting the treating staff.

From the end of January 2020, an order established the mandatory isolation in a dedicated hospital setting of any person who returned from China showing symptoms of illness. The order also stated that forced isolation can be carried out for a person who shows resistance to voluntary isolation. At the same time, flights arriving from China were prohibited from landing in Israel, and later the entry of tourists to Israel from East-Asian countries (Japan, Hong Kong, South Korea, Thailand, etc.) were prohibited. In February 2020, airline routes from East Asian countries to Israel were closed, followed by routes from countries with excess morbidity – Italy and Spain. Finally, in the second week of March, air traffic to Israel was significantly reduced to the point of almost complete closure of Israel's air, land and sea borders (with the exception of rescue flights to repatriate Israeli residents abroad). At the same time (February 2020) a mandatory 14-day home isolation period was established – first for those returning from China and other East Asian countries, then for those returning from Italy, Spain, France, Germany, Switzerland and Austria. Finally, in the second week of March, the obligation of home isolation was extended to all those entering the State of Israel.

The first COVID-19 patient was diagnosed in Israel on February 23, 2020. The turning point of disease spread occurred in the second week of March 2020, when an exponential increase in new cases was observed in the country. Between March 2020 and March 2022, Israel faced five COVID-19 waves which resulted in over 4 million cases and over 10,000 deaths.

Measures taken during the first wave

In March gatherings and mass events were gradually decreased from 2,000 to 100 participants and were then banned. Places of recreation and leisure as well as workplaces were closed (except for those defined as essential workers – 15% activity in the economy). On March 17, 2020, lockdown was declared. The education system and universities were shut down. Public transportation was significantly reduced, and residents were instructed not to leave their house, except for essential needs (food, medication and essential work). Later, the citizens were instructed not to go further than 100 meters of their place of residence (except for essential needs) and the gathering of more than two people who do not leave in the same household was prohibited. These instructions were anchored by the government as emergency regulations on March 25, 2020. In an unprecedented manner, in March 2020 the Israeli government decided to make use of advanced technological means of cellular tracking in order to enforce the obligation of isolation. In addition, and as a tool for epidemiological investigation, the General Security Service was authorized (under emergency regulations) to collect and process detailed information about the location and movement routes of people who were diagnosed with COVID-19 from 14 days before the diagnosis, with the aim of identifying contacts and isolating possible infection circles. Once the system was activated, individuals began to receive proactive messages about being in the vicinity of a confirmed COVID-19 patient without revealing the details of the patient himself. This move to track individuals by cell phones has received a lot of public criticism on the grounds of a severe and disproportionate violation of basic rights, including the right to privacy.

In the second week of April 2020, due to fear of gatherings and contagion during the Jewish holiday Passover, it was decided to tighten lockdown and move to a temporary state of curfew for the entire days of the holiday. Police and judicial forces were deployed across the country while blocking mobility between cities and strictly enforcing those who violated the curfew conditions (among other things by imposing fines between NIS 500 and NIS 5,000). Lockdown restrictions were gradually lifted in May and June 2020.

Measures taken during the second wave

On September 18, 2020, a second lockdown for 21 days was announced in Israel. A few days later, it was decided to take further steps to tighten the lockdown. A differential program called "The Traffic Light", which comprised classification of cities according to morbidity levels, was instated. On October 17, 2020, lockdown ended, except for cities defined as 'red cities' under the Traffic Light program.

Measures taken during the third wave

On December 27, 2020, due to a renewed increase in morbidity, a third lockdown was announced. Unlike the previous two lockdowns, this time the Israeli government

decided not to close educational institutions and the scope of work in the private sector was only reduced to 50%.

Measures taken during the fourth and fifth waves

From June 2021, after elections and the establishment of a new government, Israel decided to move from a policy of lockdowns in response to increased morbidity to softer preventive measures. These included encouraging vaccinations, reducing gatherings, providing green certificates to vaccinated individuals which allowed them to enter shops and other public places, and monitoring morbidity.

The use of models to predict morbidity and mortality

Reference scenarios are based on mathematical models used by public health experts to study infectious diseases. Such scenarios have many advantages, but they also present significant challenges, disadvantages and are sometimes even misleading. Starting in March 2020 experts from various fields began publishing in the public domain predictions on the "behavior" of the Severe Acute Respiratory Syndrome Corona Virus 2 (SARS-CoV-2) and the expected morbidity, mortality, and duration of the pandemic. These scenarios ranged from very optimistic to pessimistic ones. The models were designed to assist decision makers in dealing with core questions about the pandemic, such as the expected daily number of infections, the expected burden on hospitals, and policy implications. The predictions were based on mathematical calculations, hypotheses and assumptions, but on very little reliable information.

The Israeli Ministry of Health began to work with several reference scenarios formulated based on various assumptions at a relatively early stage. Most of the models were based on the Susceptible-Infectious-Recovered (SIR) model for studying epidemic spread as first published in 1927 by Kermack and McKendrick (5). Although this mathematical model has evolved and developed over the years, its basic principles have not changed. Thus, most models used for predicting COVID-19 disease spread have used the basic reproduction number (R0) as a tool to reflect the intensity of epidemic spread and have applied the principles of population classification into 3 groups: susceptible, infected and recovered. R0 is the mean number of secondary cases an infected person can cause in a population where there is no immunity. R0 was calculated by comparing the number of infected individuals in a given week with the number of infected individuals in the previous week. This index is greatly affected by a range of factors that can be influenced both by the characteristics of the population and by preventive measures. Although R0 is an important tool for developing theoretical models, its effectiveness in predicting the spread diseases was not tested prior to the COVID-19 pandemic. The most optimistic scenario (R0 = 1.2) estimated that there would be 108,000 critically ill patients with COVID-19 in Israel and that 8,600 would die of the disease, while the most pessimistic one (R0 = 2) predicted that there would be 270,000 critically ill patients and that 21,600 would die (The definition of critical cases was based on the Israeli Ministry of Health's definition, which included oxygen saturation levels below 94% as the main criteria). These scenarios were first presented to decision-makers and later to the public without proper and balanced mediation of the information, which led to the escalation in national anxiety and panic. At first the panic was translated into a rare public collaboration with government directives, but as time went on and the actual number of cases, critically ill patients and deaths turned out to be fundamentally different, the scenarios and forecasts became a double-edged sword as public trust and cooperation began to falter.

Although reference scenarios are based on a seemingly objective mathematical models, they often embody a variety of subjective assumptions due to the uncertainty inherent in unfamiliar morbidity. Models can only be useful in the context of imperfect information. In the absence of reliable information, some of these assumptions are influenced by the modelist's personal and/or professional perceptions of the characteristics of the disease, its future behavior, and the behavior of the public.

Assessment of gaps between models and real-world data, can assist policy makers in adopting an informed, data-based approach and can advance knowledge-based decision-making processes when dealing with subsequent crises. We examined the various reference scenarios that were presented to decision-makers in Israel and tried to estimate their quality compared to the actual data collected.

Study data and methods

Although many prediction models were developed during the COVID-19 pandemic, we analyzed the 3 major models that served as the basis for policy making. Model data used in this analysis is public information that was made available by the three model teams at different timepoints during the pandemic on social media, television, press, official social networks, and as presentations presented to health policy leaders.

The models were analyzed anonymously. Due to the multiplicity of existing data in the field, we focused on comparing the predicted number of confirmed COVID-19 patients, patients in critical condition and deaths to the actual data collected and published in the Israeli Ministry of Health's dashboard.

For convenience, we defined the five COVID-19 waves in Israel:

First wave: March–May 2020

Second wave: June–November 2020

Third wave: December 2020–May 2021

Fourth wave: June–November 2021

Fifth Wave: December 2021–February 2022.

For the purpose of this study, a model forecast beyond ±10% of the 'real world data' is over/under estimating.

Study results

The progress of the number of new cases in Israel is presented in Figure 1. Table 1 presents a comparison of the models and real-world data published by the Israeli Ministry of Health. It is important to note that the table is based on publications of the models and thus reflect different outcomes, which differ between the models. For example, while some of the models presented the cumulated number of new cases, others focused on the highest (peak) number of new cases and did not provided forecast of the cumulated number of cases for the entire wave.

Models description

There were inherent differences between the different models as each model uses different techniques and relies on specific characters and scientific approaches. Severe cases were defined by the models based on the MOH definition (saturation levels, intensive care units). The models evolved over time to include additional variables such as vaccination status of various populations, the effect of various stringency of lockdowns, and other preventive actions. The purpose of team number 1 was to model the risk for the collapse of the health system. This was achieved by modeling the chance for hospitalization and death for each new case. The team developed a new model which was used with Monte Carlo simulation and combined models based on the characteristics of the infected individuals. The model included a survival analysis with Kaplan Meier and a Cox proportional hazards model. Team number 2 used an Age-of-Infection model. It uses the number of cases which has characteristics of Poisson distribution with dynamic expectation. Team number 3 performed short-term modeling (nowcasting) combined with other stochastic models (agent-based models) that examined the individual and followed the course of the exposure (infected, hospitalized, deceased, etc). The model presented a forecast of the presence and the near future and aimed to identify important parameters and provide predictions based on these parameters.

The first wave: March–May 2020

The models' overestimation of morbidity in comparison to actual data is highly evident in the analysis of First-Wave data. The predicted cumulative number of confirmed cases for this wave was up to 100 times higher than the actual number of confirmed cases. A similar trend was observed for the comparison between the predicted and actual number of deaths and critically ill patients. The largest gap between predicted and actual data was observed in Team 3's model. Team 3 predicted that the number of confirmed cases would be 50–100 times higher than actual data. Interestingly, Teams 2 and 3 predicted the same range of deaths, probably because they used R0 as the main explanatory variable in their sensitivity analyses. Team 1's number of predicted deaths was 161 times higher than the actual data (46,772 vs. 289).

The second wave: June–November 2020

Ministry of Health data showed that at the Second Wave's peak, at the end of September 2020, there were 9,051 confirmed cases and the cumulative number of confirmed cases for the entire wave was 319,921. This number corresponds with the models' predictions. However, all three models provided a significant overestimation of the number of critically ill patients (495 at the wave's peak). Teams 1 and 3 overestimated this number by 2–3 times. Team 2's estimation was even higher, but this team combined the number of patients in moderate and critical condition. Interestingly, Teams 2 and 3 predicted a significantly lower number of deaths compared to the actual data, while Team 1 overestimated the number of deaths by more than 100%. It is important to note that Team 1 only predicted the number of new patients in critical condition and did not provide predictions for the number of confirmed cases. Therefore, it is difficult to assess the effectiveness of the model in predicting disease spread (in terms of new cases) during the second wave.

The third wave: December 2020–May 2021

In the Third Wave Team 2 provided the closest prediction to actual data – both in the number of confirmed cases and in the number of deaths. This team tried to evaluate the effectiveness of strict lockdown alongside the vaccination campaign; therefore, it made a relatively careful estimation (under estimation) of the number of confirmed cases. Actual data showed an initial increase in the number of confirmed cases which decreased slowly. This trend was not predicted by the model. The models that were relatively accurate in predicting the number of confirmed cases in the second wave, made significant overestimations in the third wave, probably due to the start of the vaccination campaign.

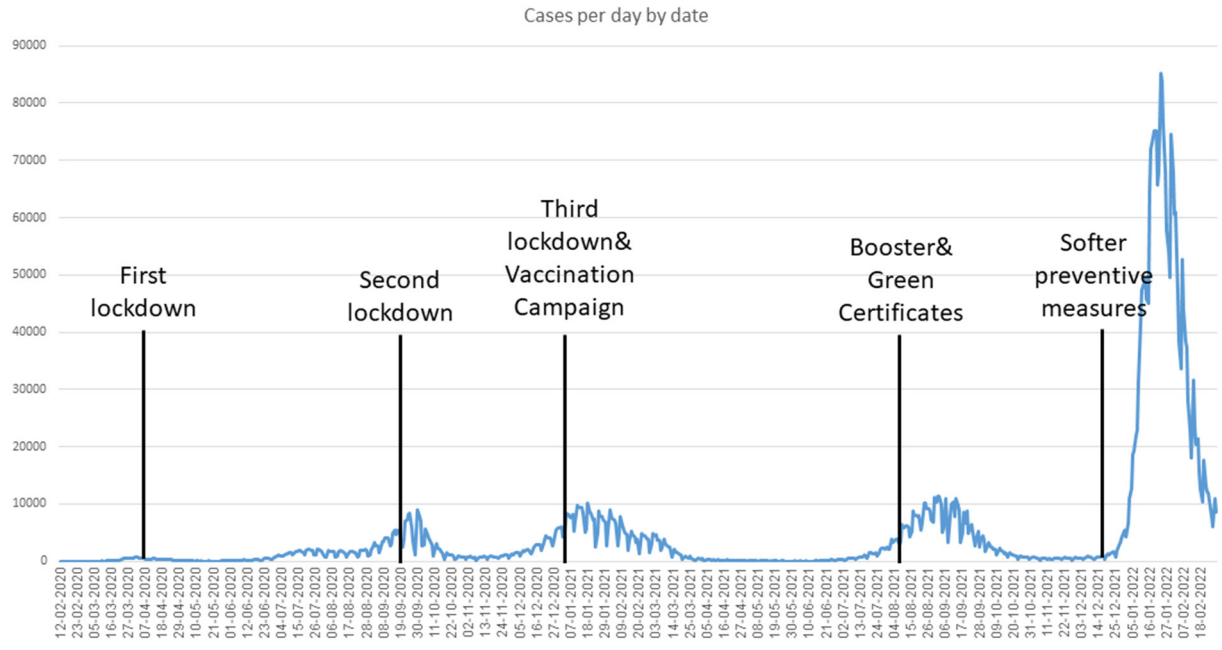


FIGURE 1
Number of confirmed COVID-19 infections per day in Israel (March 2020–February 2022).

The fourth wave: June–November 2021

The fourth wave was characterized by the spread of the delta strain of the virus, along with a decline in immune defense. Most teams were successful in predicting the number of confirmed cases as well as the number of deaths, with only about 10% deviation between the model and actual data. In contrast, the models, and particularly that of Team 3, overestimated the number of critically ill patients (2000 predicted vs. 766 actual). Team 2 developed a model which considered the vaccination booster as a crucial parameter in predicting the number of critically ill (1000–2500). However, both predicted values presents a large-scale overestimation. It is possible that the gradual decline in immune protection may have benefited the models by reducing the gap between the actual number of critically ill patients and their predicted number. Although the models did not account for decreased immune protection but rather provided their prediction based on the number of confirmed cases, the decrease in immune protection led to an increase in the number of critically ill patients, bringing it closer to the predicted number.

The fifth wave: December 2021–February 2022

The rapid spread of the omicron variant, which manifested in a very high number of daily confirmed cases (up to a peak of 85,000 confirmed cases per day and a total of about 2.3 million confirmed cases in this wave), resulted in high variance among the models regarding the number of confirmed cases and the rate of increase in cases. While Team 1 predicted an average of 45,000 confirmed cases per day, Teams 2 and 3 predicted a cumulative number of confirmed cases that was double compared to the actual data: Team 2 estimated that there would be 4 million confirmed cases (vs. 2.3 million cases that were actually confirmed) and Team 3 predicted that there would be 73,000–146,000 confirmed cases per day. There was considerable variation among the models in the predicted number of critically ill patients: while Team 1 predicted a lower number of critically ill patients than the actual number, Team 2 overestimated their number by over 100%, and Team 3, which provided gloomy predictions for most of the waves, accurately predicted the number of critically ill patients in the fifth wave. It is also important to note that due to multiple contradictory information on the severity of disease caused by the Omicron strain, information on the predicted number of deaths was not published by the modeling team. Instead, they chose to focus on the number of newly confirmed cases and the number of critically ill patients.

Discussion

Models' evaluation

From the day it was identified, SARS-CoV-2 has proven its ability to surprise healthcare systems worldwide. Tthroughout the COVID-19 pandemic in Israel, the reference models

TABLE 1 Comparison of models and real-world data, Israel March 2020–February 2022.

	Ministry of health data	Team 1	Team 2	Team 3
Model description		Combined models based on the infection characteristics: Survival analysis Kaplan Meier) and Cox proportional hazards model Aalen johsnaen estimator self-developed model with Monte Carlo simulation	Age-of-Infection model: number of cases has Poisson distribution characteristics with dynamic expectation	Short-term modeling (nowcasting) combined with other stochastic models (agent-based models)
First wave: March-May 2020				
Confirmed COVID-19 cases Tests	Peak: 17,047 (peak 724/day) No data	–	18,000–193,000	576,000–1,440,000
Critically ill	192	224,366	108,000–270,000	108,000–270,000
Dead	289	46,772	8,600–21,600*	8,600–21,600*
Second wave: June—November 2020				
Confirmed COVID–19 cases Tests	319,921 (Peak 9,051/day) 3,362,484		298,000	7,500–9,200 peak
Critically ill	495	976–1,193 Weekly forecast	critically and moderately ill: 9,550	820
Dead	2,596	From March 20: 8,361–10,534 (8,000–10,200 for the second wave alone)	1300	, 1,600
Third wave: December 2020—May 2021				
Confirmed COVID–19 cases Tests	Peak: 10,123 502819 (peak 10,123/day) 10,375,126	16,700–22,300 Depending on sensitivity analysis for vaccine efficiency and initial R	Differs between different scenarios (strict lockdown yes/no) 6,120–10,000	4,000–8,000 peak
Critically ill	1,193	1,700–2,400	Differs between different scenarios (strict lockdown yes/no) and on vaccination status 600–1,474 (lockdown y/n) 1,340–3,230 (vaccine y/n)	2,539–6,834 (vaccine y/n)
Dead	3,541	2,450–2,700	Differs between different scenarios (strict lockdown yes/no) and on vaccination status 1,250–3,085 (lockdown y/n)	4,500–5,700 On average, depending on preventive actions
Fourth wave: June—November 2021				
Confirmed COVID–19 cases Tests	50,4587 (Peak 11,346/day) 18,437,810	11,000 peak	Mainly children and unvaccinated adolescents	9,000 peak
Critically ill	766	850	Varied based on the vaccination status and based on age. 1000 (with booster)–2,500 (no booster)	2,000
Dead	1,782	–	200–1,400 Depends on the rate of the vaccination campaign (combined model teams 2 and 3)	1,500 according to mortality model
Fifth wave: December 2021—February 2022				
Confirmed COVID–19 cases Tests **	2,293,405 (Peak 85,192/day) 18,369,900	45,000/day	4,000,000	73,000–46,000/day
Critically ill	1,255	800–1,000	1,250–2,750	700–1,566
Dead	2,042			100/week

*Teams 2 and 3 worked together at the beginning of the pandemic, joining forces in order to provide scientific assumptions to policy makers. As a result, the same prediction was provided by both teams. Later on, as these teams separated, differences in forecasts were observed. **The number of tests during the fifth wave includes both PCR tests and Rapid (Antigen) tests.

10.3389/fpubh.2022.1002440

overestimated morbidity in comparison to actual data from the Ministry of Health. High variance was observed both among the models and for each model over time. Each modeling team chose to focus on a different outcome. For example, in later waves Team 1 moved to focus on patients in critical condition while Team 3 continued to provide predictions regarding the number of confirmed COVID-19 cases.

An examination of the various models by time shows that in most cases all teams overestimated the number of confirmed COVID-19 cases and critically ill patients with better accuracy in later stages of the pandemic. This is probably a consequence of the lack of uniformity in the knowledge available to the various modelers during the pandemic. The models usually used the R0 number as the primary variable for predicting the spread of disease. Considering the evolution and variability of SARS-CoV-2 during the pandemic, this classification proved to be challenging. The uncertainly derived from the lack of reliable and available information on a wide range of essential parameters that affect R0, including, the nature of infection, its duration and severity, transmissibility, infectiousness, the average number of days in which an individual remains infected, population density and health, the average age of the population, the appearance of new variants, and the behavior of the public. In addition, the published Rt and the number of those infected were affected by government measures, the extent of public immunization, the effect of vaccines on the number of new cases and on the number of critically ill patients and the degree of protection against re-infection. Furthermore, the various models examined disease spread on a national level and did not make consider essential variables and unique characteristics to Israel that may influence disease spread, such as its young population, one central entry into the country, emergency preparedness, the population strata, its density, and the number of children per household. Moreover, due to the great complexity of the spread characteristics of the virus, it was not possible to construct a model that included all known variables and the researchers were satisfied with relatively simple models that included using R0 numbers from other countries or calculating R0 numbers according to the number of confirmed infections in Israel with age adjustments. All of these variables created a complex reality that challenged the various models (6). In practice methods that previously helped in predicting disease spread have been found to be less accurate, (7, 8) which has led, among other things, to constant updating of models and methods (9, 10). Over time, each modeling team chose to analyze different data. Although they could better predict the wave trend (increase/decrease), wide gaps remained between the data presented by each model and the actual intensity of morbidity, the number of critically ill patients and mortality. Models that provided accurate predictions for one wave, were very inaccurate in the next wave.

Public communication

The public's behavior also affected the spread of disease: during the first wave there was great uncertainty together with conflicting reports from sources abroad about a very high R0 number, which contributed to very high public response to the restrictions imposed by the government. As time passed and the public understood that the pandemic may last for a long time, compliance with governmental restrictions decreased, which may have resulted in certain gaps between the models and actual data. For example, while the models predicted a steep rise in morbidity, no such increase was recorded in real life. These gaps may have led to further public non-compliance with governmental restrictions because the predictions did not come true.

The differences between the models' predictions and real-world data shows that policy-making cannot stop the spread of disease altogether, however, it has a limited ability to slow it down, with the intention of trying to flatten its growth curve as much as possible and to avoid very high morbidity in a very short time (and thus to avoid insufficiency of the health system), sometimes even at the cost of dispersing disease spread over a longer period of time.

It is also important to consider the issue of disseminating the models to the public. The extensive media coverage of the COVID-19 pandemic included, among other things, the daily publication of morbidity data alongside various assessments, forecasts and models. Often, this information was partial or was provided without framing the information within the correct context, so that the public only received the "bottom line" (R0 number scenario or prediction of morbidity) without the different parameters that make up the model, the sensitivity tests and parameters that affect the model's accuracy in predicting morbidity, or the difference in the meaning of the scenario (advantages/disadvantages) and its predictions. Due to the recognized importance of models as a tool to help dealing with the pandemic, and the understanding that in the reality of a new disease much more is unknown, it would have been more appropriate to mediate the information in a way that reflected its limitations and shortcomings and to show professional modesty.

Although transparency is a fundamental value in health systems, the damage to public trust that results from publishing information without appropriate mediation and context framing outweighs its benefit. Such damage to trust can lead to decreased public response to professional guidance and to the ability of the government to deal with the spread of disease. In practice, the public's exposure to significant gaps among the scenarios predicted by the models and actual morbidity, increased the public's distrust in decision-makers. Hence, it is very important to mediate the information and the various models to the public in an orderly and reliable manner, while presenting the models correctly and accurately as a tool for decision-making.

Advice/policy implications

Upon the emergence of a new disease, there is an inherent difficulty in relying on mathematical models as a reliable tool for predicting and formulating a strategy for dealing with its spread. Such uncertainty and a multiplicity of assumptions that are not based on reliable information may lead to significant gaps among the various models, and between the models and real-world data. Data researchers who have agreed to contribute their time and experience toward presenting ways to deal with the COVID-19 pandemic are a welcome phenomenon that should be encouraged and preserved. At the same time, decision-makers must integrate the information presented to them in order to advance knowledge-based decision-making processes on the one hand, but they must also recognize the structural weaknesses of mathematical models when faced with uncertainty. The decision-making process and health policy design should regard models as auxiliary tools and consider their limitations and weaknesses, while remembering that preventive measures, public behavior, seasonality and additional factors influence the models' predictions. Behavioral elements such as public compliance, avoiding crowding, participating in indoor activities, etc. has a large impact on the different models accuracy. Thus, real world data on these parameters can be integrated into the models in order to enhance their precision.

It is essential to correctly mediate the reference scenarios to decision makers and the public, while providing the appropriate context of the mathematical models together with their advantages and disadvantages. In view of the findings of this study, we suggest creating an elaborate mechanism that will serve as a tool for decision-makers. This mechanism should comprise two separate but complementary components: (a) a prediction range derived from combined key models; (b) an independent and separate prediction for each of the models while preserving their different methodologies. Furthermore, we recommend developing a mechanism that would provide modelers access to institutional data in a structured and orderly manner, in addition to the information collected by them independently. This may help to improve and refine the various mathematical models.

Limitations

First, the current study presents data from different models that were used by policymakers. These models were influenced by several parameters such as policy recommendations, preventative measures, and public awareness, social distancing, etc. this may lead to a potential bias in providing interpretations for the results. However, policy makers used the same models, with the same potential biases, thus we believe the analysis suggested is appropriate. Second, the models were published only in secondary publications and were not peer-reviewed. Nevertheless, these models were presented to the government and health authorities and they served as the basis for decision making.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://datadashboard.health.gov.il/COVID-19/general.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

The authors would like to thank Mr. Eldad Sitbon and Oliver Geffen for their insights and fruitful brainstorming that were useful in improving this study.

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.

Publisher's note

References

1. Colizza V, Barrat A, Barthelemy M, Valleron AJ, Vespignani A. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS Med.* (2007) 4:e13. doi: 10.1371/journal.pmed.0040013

2. Dye C, Gay N. Epidemiology modeling the SARS epidemic. *Science.* (2003) 300:1884–5. doi: 10.1126/science.1086925

3. Hufnagel L, Brockmann D, Geisel T. Forecast and control of epidemics in a globalized world. *Proc Natl Acad Sci U S A.* (2004) 101:15124–9. doi: 10.1073/pnas.0308344101

4. Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. *Acta Biomed.* (2020) 91:157–6. doi: 10.23750/abm.v91i1.9397

5. Kermack WO, McKendrick AG, Walker GT. A contribution to the mathematical theory of epidemics. *Proc. Royal Soc. Contain. Papers Mathematic. Phys. Character.* (1927) 115:700–21. doi: 10.1098/rspa.1927.0118

6. Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic? *Infect Dis Model.* (2020) 5:271–8. doi: 10.1016/j.idm.2020.03.001

7. Longini IM. A mathematical model for predicting the geographic spread of new infectious agents. *Math Biosci.* (1988) 90:367–83. doi: 10.1016/0025-5564(88)90075-2

8. Mathews JD, McCaw CT, McVernon J, McBryde ES, McCaw JM. A. biological model for influenza transmission: pandemic planning implications of asymptomatic infection and immunity. *PLoS ONE.* (2007) 2:e1220. doi: 10.1371/journal.pone.0001220

9. Tang B, Bragazzi NL Li Q, Tang S, Xiao Y, Wu J. An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). *Infect Dis Model.* (2020) 5:248–55. doi: 10.1016/j.idm.2020.02.001

10. Li G, Chen K, Yang H, A. new hybrid prediction model of cumulative COVID-19 confirmed data. *Process Saf Environ Prot.* (2022) 157:1–19. doi: 10.1016/j.psep.2021.10.047

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 19 December 2022
DOI 10.3389/fpubh.2022.1037818

OPEN ACCESS

EDITED BY
Pierpaolo Ferrante,
National Institute for Insurance Against Accidents at Work (INAIL), Italy

REVIEWED BY
Charles J. Vukotich Jr,
University of Pittsburgh, United States
Igor Nesteruk,
National Academy of Sciences of Ukraine, Ukraine

*CORRESPONDENCE
Tong Dong
dongtong.2022@163.com
Ying Wang
1809231295@qq.com

SPECIALTY SECTION
This article was submitted to Public Health Policy, a section of the journal Frontiers in Public Health

RECEIVED 06 September 2022
ACCEPTED 28 November 2022
PUBLISHED 19 December 2022

CITATION
Yin S, Ma L, Dong T and Wang Y (2022) Measuring the impact of the COVID-19 epidemic on university resumption and suggestions for countermeasures.
Front. Public Health 10:1037818.
doi: 10.3389/fpubh.2022.1037818

Measuring the impact of the COVID-19 epidemic on university resumption and suggestions for countermeasures

Shi Yin[1], Lijun Ma[2], Tong Dong[1]* and Ying Wang[1]*

[1]College of Economics and Management, Hebei Agricultural University, Baoding, China, [2]College of Land and Resources, Hebei Agricultural University, Baoding, China

Background: During the COVID-19 pandemic, universities around the world had to find a balance between the need to resume classes and prevent the spread of the virus by ensuring the health of students. The purpose of our study was to effectively assess the overall risk of universities reopening during the COVID-19 epidemic.

Design and methods: Using the pressure–state–response model, we designed a risk evaluation method from a disaster management perspective. First, we performed a literature review to find the main factors affecting the virus spread. Second, we used the pressure–state–response to represent how the considered hazards acts and interacts before grouping them as disaster and vulnerability factors. Third, we assigned to all factors a risk function ranging from 1 to 4. Fourth, we modeled the risk indexes of disaster and of system vulnerability through simple and appropriate weights and combined them in an overall risk for the university resumption. Finally, we showed how the method works by evaluating the reopening of the Hebei Province University in 2022 and highlighted the resulting advice for reducing related risks.

Results: Our model included 20 risk factors, six representing exogenous hazards (disaster factors) that university can only monitor and 14 related to system vulnerability that can also control. Disaster factors included epidemic risk level of students' residence and the school's location, means of transportation back to school, size of the university population, the number of migrants on and off campus and express carrier infection. Vulnerability factors included student behaviors, routine campus activities and all the other actions the university can take to control the virus spread. The university of Baoding city (Hebei Province) showed a disaster risk of 1.880 and a vulnerability of 1.666 which combined provided a low risk of school resumption.

Conclusion: Our study judged the risks involved in resuming school and put forward specific countermeasures for reducing the risk levels. This not only protects public health security but also has some practical implications for improving the evaluation and rational decision-making abilities of all parties.

KEYWORDS

resumption of universities, risk prevention and control, comprehensive risk assessment, COVID-19, fuzzy statistical model

1. Introduction

The novel coronavirus disease 2019 (COVID-19) is an ongoing pandemic that has evolved into a global crisis and has seriously challenged the development process of human society (1). The COVID-19 pandemic is a major worldwide public health emergency that spreads fast over a wide range of locations and is difficult to prevent and control; however, thanks to the concerted efforts of people around the world and to a global vaccination campaign, the epidemic prevention and control situation has continued to improve, and the order of work and life was quickly restored (2). Despite this, the appearance of virus variants with higher infectivity makes the pandemic still not under effective control (3). Tertiary education Institutions play a key role in assuming the functions of higher education and are also an important public field to manage emergencies (4). In the COVID-19 context, many colleges and universities have issued plans to reopen their school even if before the pandemic showed high incidences of respiratory infectious diseases. While resuming in-person teaching, the back-to-school activities mean further battles for epidemic prevention and control for several reasons. First, colleges and universities need to be effective regarding the prevention, monitoring, and management of public health emergencies. Measures taken by colleges and universities often lag behind the development speed of the crisis, and the phenomenon of post-management rather than prevention always exists. For example, emergency warning mechanisms in Chinese universities have been ineffective due to the uncertainty of public health emergencies and technology defects. Second, the ability to cope with and guide public opinion in a public health crisis is a considerable aspect of university governance modernization, especially in the Internet era. As a relatively closed social cluster, the spread of rumors and false information in a university can easily cause panic among teachers and students, creating additional considerations for governing public health emergencies (5). Social media has rapidly developed, and society has entered an information age. Young people use social media much more frequently than other age groups, making it more difficult for universities, which are mainly composed of young college students, to curb false information. Third, the education system has entered the mobile war stage of epidemic prevention and control, and risk factors have become more complex and changeable (6). The full resumption of education in colleges and universities has brought about a larger range and scale of personnel mobility. The activities and management issues of overseas students have brought new risks to epidemic prevention and control in schools, which has changed from positional to mobile anti-disease warfare. In addition, as epidemic prevention and control has entered the normalization stage, various associated problems emerged, such as stress responses, anxiety, and other psychological problems; livelihood issues, such as entering schools and resuming employment; and teaching management issues, such as the connection between online and offline teaching. In China, the problem has been investigated with different approaches and several solutions have been proposed. Yang (7) constructed a risk assessment system for school respiratory infectious disease outbreaks from four perspectives: possibility, vulnerability, severity, and countermeasures. Liu and Zhang (8) discussed applying a risk assessment of the overall smart campus framework in terms of risk identification, assessment, disposal, and control to form a set of network security risk assessment methods that can be widely applied to the current overall smart campus frameworks. Ding and Li (9) proposed a Delphi method and AHP method combined with Borda ordinal value method to study the risks after returning to school under the COVID-19 epidemic. Wang et al. (10) discussed a risk assessment method for reopening universities that can evaluate the comprehensive risks of resuming education during the COVID-19 epidemic and assist universities in making organizational decisions for reopening. Although the latter study analyzed the interaction mechanisms of various factors based on pressure–state–response model and established a comprehensive risk assessment index system for COVID-19 outbreaks in colleges and universities, important factors involved during education resumption are missing. Starting from the previous results, the aim of this study is to introduce a comprehensive index to measure the risk of virus spread during university resumption and to take a university as our research object.

2. Methods

The proposed risk evaluation tool is designed from a disaster management perspective and is based on the pressure–state–response model. First, we have selected the main factors affecting the virus spread that are the most considered in literature. The selection of these factors is based on the five principles of significance, operability, practicability, relevance and concreteness of the index system, PSR model, and many literature (2–22). Second, we used the pressure–state–response to represent how those hazards acts and interacts before grouping them in disaster and vulnerability factors. Levels of categorical factors (such as means of transport) were ranked from lowly to highly dangerous and recoded with values corresponding to the risk rank (such as Self-driving = 1, Taxi = 2, Train/R = 3, Other = 4). Third, according to Chinese 3-levels territorial epidemic risk (high, medium, and low),[1] to simplify the computational process we used step functions to assign risks

1 In June 2022, the following classification was introduced: Over the past 14 days, areas with no confirmed cases nor new cases are defined as low risk; areas with no more than 50 newly confirmed cases or a cumulative total of more than 50 confirmed cases and no cluster of outbreaks are defined as moderate risk. Areas with more than 50 cumulative confirmed cases and clusters of outbreaks are defined as high risk.

to factors values ranging from 1 to 4. The choices of those functions are based on the experience of the risk assessment expert group composed of experts in various fields such as medicine, education and emergency management. Fourth, we modeled the risk indexes of disaster and of system vulnerability as weighted mean of related indexes and combined them in and overall risk for the university resumption. Finally, we showed how the method works by evaluating the reopening of the Hebei Province University in 2022 and highlighted the resulting advice for reducing related risks.

2.1. PSR model

The PSR framework models the chain of causal links between a system working to maintain a state, and exogenous forces working to change it. Formally, it is an interconnected conceptual structure consisting of three parts (11): pressure, which represents the process of adverse effects generated from the system interference and coercion (12); state, which represents the current state of the system under external pressure (13); response, which represents the feedback process of the system in response to external pressure (14). Even if domestic scholars generally use this theory to explain phenomena in the fields of taxation and ecological and environmental protection (15), it can also reflect the dynamic processes and internal logic of university environments. All the factors that affect the epidemiological risks associated with university resumption and all the involved subjects interact and influence each other in a dynamic balance (Figure 1). The epidemic situation before students return to school, the public transportation they take on the way back to school, and the flow of people inside and outside the school after they return increase the risk of epidemic transmission and together form the pressure system. Information released by the pressure system allow involved subjects to take countermeasures. The status system includes student behaviors and routine campus activities, such as raising students' risk awareness, adopting online education, distance learning and strengthening campus space management to reduce the risk of the virus spreading. The response system includes all the measures a university can take that do not strictly concern the routine campus activity, such as emergency plans, drills, and assessments; the establishment of isolation sites; and the development of response systems to deal with outbreaks. The epidemic prevention and control status quo in colleges and universities will directly affect the response measure effectiveness and timelines, and improvements in response capacity will feed back to the status system and improve the prevention and control status quo.

Methods for detecting and measuring selected risk factors are summarized in Supplementary material 1 and described in the following sections.

2.1.1. Indicators of pressure system

Since the movement of people inside and outside the school will increase the risk of the epidemic spreading, as indicators of the Pressure System we selected: epidemic risk level of students' residence (P_1) (13); epidemic risk level of the school's location (P_2) (6); means of transportation back to school (P_3) (6); size of the university population (P_4) (16); the number of migrants on and off campus (P_5) (16); and express carrier infection (P_6) (17). According to national classification of territorial risk (see text footnote 1), we assigned to each level of P_1 and P_2 risks equal to 1.5 (low), 2.5 (medium) or 3.5 (dangerous). Risk of P_3 was assessed as the weighted mean of risks of the 4 most used vehicles such as Self-driving, Taxi, Train/R, Other. Means of transportation were ranked by the expected number of close contacts, then we assigned equidistant risk values (Self-driving $= 1$, Taxi $= 2$, Train/R $= 3$, Other $= 4$) weighted with fraction of users. Risk values of 1, 2, 3, 4 were assigned to P_4, P_5, and P_6 as follows: if the number (n) of students returning to school/1,000 was $n < 0.5$, $0.5 \leq n < 1$, $1 \leq n < 2$, $2 \leq n$ respectively; if the percentage (p) of students entering and leaving the school was $p < 5\%$, $5\% \leq p < 15\%$, $15\% \leq p < 25\%$, $p \leq 25\%$ respectively; if the percentage (p) of express deliveries from medium and dangerous risk places was $p < 5\%$, $5\% \leq p < 15\%$, $15\% \leq p < 25\%$, $p \leq 25\%$ respectively.

2.1.2. Indicators of state system

We analyzed the state system from the perspective of individual students, the campus environment, management measures, and the use of digital technology. As indicators we selected: students' knowledge of the epidemic (S_1) (9); students' awareness of risk (S_2) (10); the situation of campus environment renovation (S_3) (18); propaganda of anti-epidemic culture (S_4) (19); personnel control measures (S_5) (20); periodic nucleic acid testing (S_6) (21); campus space management (S_7) (18); the application of digital teaching technology (S_8) (22); and the application of digital epidemic prevention technology (S_9) (6). Risk of S_1 was assessed by assign risk values to the students' knowledge level of epidemic prevention and control (X) of $4 - 3(X - 8)/(10 - 8)$ and 4 if $8 \leq X$ and $X < 8$, respectively. Risk values of 1, 2, 3, 4 were assigned to S_2–S_9 as follows: if the risk awareness of the students observed in daily behavior was very strong, strong, general, and weak, respectively; if the campus environment renovation (cleaned up campus health dead spots; disinfected public places; equipped with disinfectants and hand sanitizer) was completed (three items), almost completed (two items), started (one item), and not started (0 item), respectively;

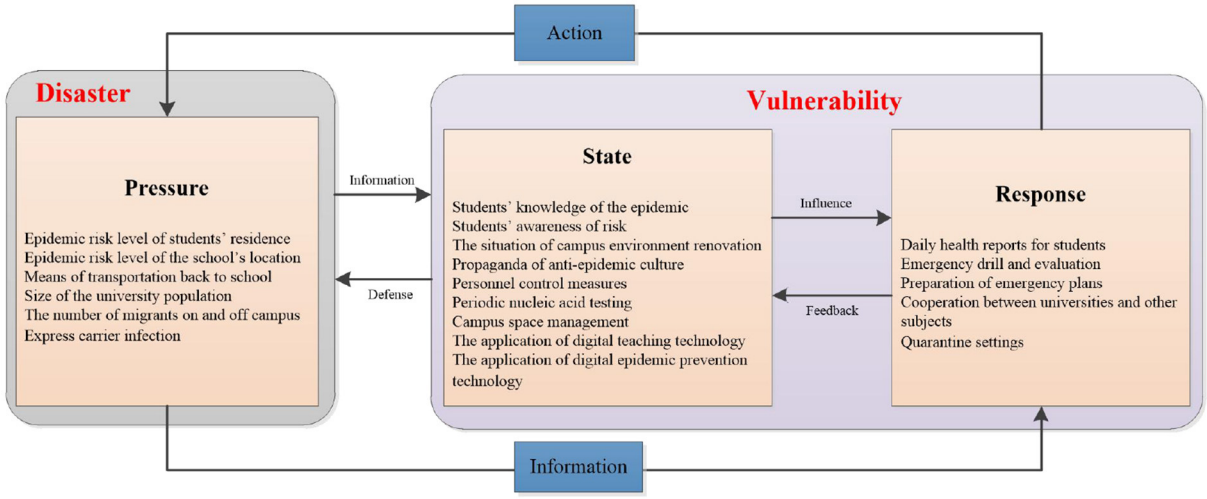


FIGURE 1
PSR mechanism model of epidemic risk situation in universities from a disaster management perspective.

TABLE 1 Value reference for r_i .

w_i^*	Instruction
1.0	w_{i-1}^* and w_i^* have the same contribution
1.2	w_{i-1}^* contributes a little more than w_i^*
1.4	w_{i-1}^* makes a bigger contribution than w_i^*
1.6	w_{i-1}^* has a stronger contribution than w_i^*
1.8	w_{i-1}^* definitely contributes more than w_i^*

TABLE 2 Epidemic risk matrix of resumption in universities.

	H		
V	**(1,2]**	**(2,3]**	**(3,4]**
(1,2]	*L*	*L*	*M*
(2,3]	*L*	*M*	*D*
(3,4]	*M*	*D*	*D*

if the percentage (p) of promoted prevention and control culture was 50% ≤ p, 30% ≤ p < 50%, 15% ≤ p < 30%, p < 15% respectively; if the organizational framework of colleges and universities was met the requirements, slightly defective, major defects, failure to formulate a reasonable organizational institutional framework, respectively; if the frequency (p) of nucleic acid testing was once a day, three times a week, twice a week, once a week respectively; if the distance (d) in meters (m) among the students was 2 m ≤ d, 1.5 m ≤ d < 2 m, 1 m ≤ d < 1.5 m, d < 1 m respectively; if the percentage (p) of the satisfaction of teachers and students with online teaching was 95% ≤ p, 90% ≤ p < 95%, 80% ≤ p < 90%, p < 80% respectively; if the percentage (p) of student nucleic acid testing system registration was 95% ≤ p, 90% ≤ p < 95%, 80% ≤ p < 90%, p < 80% respectively.

2.1.3. Indicators of response system

According to the mechanisms of the discovery, initiation, and control of emergency responses, as indicators of the Response System we selected: daily health reports for students (R_1) (6); emergency drill and evaluation (R_2) (19); preparation of emergency plans (R_3) (2); cooperation between universities and other subjects (R_4) (6); and quarantine settings (R_5) (2). Risk of values of 1, 2, 3, 4 were assigned to R_1–R_5 as follows: if the health monitoring days per student in the data system (d) was 14 ≤ d, 7 ≤ d < 14, 0 ≤ d < 7, unestablished system respectively; if the school epidemic drill situation was conducted as emergency drills for evaluation and made improvements based on evaluation comments, emergency drills and assessments but did not improve all assessments, emergency drills but did not conduct drill evaluations or unconducted respectively; if the emergency response plan was well-prepared contingency plans for various emergencies, comparatively perfect preparation of various contingency plans, inadequate preparation of contingency plans for various emergencies, failure to prepare for various emergencies respectively; if the university collaborated with three, two, one or no organizations respectively; if the isolation area and the personnel on duty was evaluated as adequate quarantine areas and staff on full-day duty, inadequate quarantine areas and staff on full-day duty, adequate quarantine areas and staff not on duty all

TABLE 3 Epidemic risk index collection—pressure system for university resumption.

Subsystem	Pressure					
Indicator	P_1	P_2	P_3	P_4	P_5	P_6
Result	Low risk	Low risk	3.32	$1 \leq n < 2$	$x_i < 5\%$	$x_i < 5\%$
Assignment	1.5	1.5	3.32	3	1	1
Weight	0.15	0.23	0.19	0.22	0.12	0.09

TABLE 4 Epidemic risk index collection—status system for university resumption.

Subsystem	State								
Indicator	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
Result	8.4	Strong risk awareness	Completed 3 items	$x_i \geq 50\%$	Slightly flawed	3 times a week	$1.5 \leq x_i \leq 2$	$90\% \leq x_i \leq 95\%$	$80\% \leq x_i \leq 90\%$
Assignment	3.4	2	1	1	2	2	2	2	3
Weight	0.06	0.09	0.16	0.05	0.12	0.17	0.15	0.11	0.09

TABLE 5 Epidemic risk index collection—response system for university resumption.

Subsystem	Response				
Indicator	R_1	R_2	R_3	R_4	R_5
Result	14	Carried out emergency drills and improved them	The preplan preparation was relatively perfect	Cooperated with two institutions	The isolation area was sufficient and the personnel were on duty all day
Assignment	1	1	2	2	1
Weight	0.16	0.22	0.20	0.18	0.24

day, inadequate quarantine areas and staff not on duty all day respectively.

2.2. Overall risk evaluation

From the previous three indicators, we obtained an overall risk evaluation by using a disaster management perspective. In the specific, the risk degree (R) of reopening a university was evaluated through the cartesian product of disaster (H) and vulnerability (V) factors (10).

$$R = H \times V \tag{1}$$

where H reflects the pressure subsystem and V both the state and response subsystems of the related PSR model (Figure 1). Like in natural disasters, the Equation (1) is suitable to represent the levels of risks related to the virus spread. Indeed, at the onset of outbreaks, interventions on hazards included in H may not be rapid enough and differences in the level of epidemic are determined by the capacity to reduce students' vulnerability. Furthermore, to consider the specific universities conditions we introduced index weight settings to adapt to local conditions. The proposed index method first needs to assess H and V as follows

$$\begin{cases} H = \sum_{i=1}^{6} p_i w_i \\ V = W_S \sum_{j=1}^{9} s_j w_j + W_R \sum_{l=1}^{5} r_l w_l \end{cases} \tag{2}$$

where p_i, s_j, r_l and w_i, w_j, w_l represent values and weights of each indicator of pressure ($i = 1, 2, \cdots, 6$), state ($j = 1, 2, \cdots, 9$), and response ($l = 1, 2, \cdots, 5$), respectively. W_S and W_R are the weights of the state and response systems, respectively, which represent the vulnerability of the students to the pandemic. To simplify the complexity of the evaluation system, we used the improved order relation method to determine the weights and to satisfy the weak consistency of the indicator (23).

Risk factors can be ranked from the most to the less important ($C_1, C_2..., C_m$) by associating to them a corresponding system of non-increasing weights ($w_1^* \geq w_2^* \geq ... \geq w_m^*$) with $\sum_{i=1}^{m} w_i^* = 1$. By using C_i to represent the indicators in the subsystem *P*, *S* and *R*, the weight calculation method of each indicator was as follows:

① Experts judged the influence importance of a subsystem's risk value according to each indicator in the subsystem and provided the weight order, which was denoted as

$$w_1^* \geq w_2^* \geq ... \geq w_m^* \tag{3}$$

② We compared the sorted weight of indicators $C_{i-1}(w_{i-1}^*)$ and $C_i(w_{i-1}^*)$, which were denoted using the following formula:

$$r_i = \frac{w_{i-1}^*}{w_i^*}, i = 2, 3, \ldots, m \tag{4}$$

For the value of r_i, please refer to Table 1 (23).

③ Weight w_m^* and w_i^* were calculated one after the other as follows:

$$w_m^* = \left(1 + \sum_{i=2}^{m} \prod_{k=i}^{m} r_k\right)^{-1}, \qquad w_{i-1}^* = w_m^* \prod_{k=i}^{m} r_k \quad i = 2, \cdots, m. \tag{5}$$

Finally, we used the risk matrix in Zhao and Wang (24) to determine the comprehensive risk level of the epidemic situation during university resumption, and the evaluation results were represented using D (dangerous), M (moderate), and L (low). Table 2 shows the risk matrix of the epidemic situation during university resumption. Risk level D indicates that students' return to school is unacceptable, and the school should immediately stop the return process and make corrections. M means that it is not expected to happen, and management decisions are made to prevent the development of risks. L means it is acceptable, and the risk control measures should be improved accordingly.

2.3. Case study

We used the university of Baoding city, Hebei Province, as case study. With permission from the Department of Education, the school resumed all in-person activities (under closed management) in February 2022. Since March 14, 2022, online teaching was promptly adopted, (with teachers teaching online at home and students choosing quiet places) because of outbreaks in all of China's provinces. Data related to the risk indicator factors were collected as follows. Risk levels issued by the regional Health Commission were used to measure P_1 and P_2. The university has performed a survey of returning students through questionnaire to measure P_3, P_4, and P_5. We asked that express deliveries station fill out the online shared form questionnaire to measure P_6. We randomly selected 100 students to conduct an epidemic knowledge questionnaire to collect data related to factors S_1 and S_8. We asked that the monitor of each class fills out the online shared form questionnaire to collect data related to factor S_2. We asked that the head of epidemic control checks the item and fills out the online shared form questionnaire to collect data related to factors S_3, S_6, S_7, and S_9. We asked that the publicity department of the school count the number and methods of anti-epidemic activities and fills in questionnaires to collect data related to factor S_4. Risk level can be determined by the risk assessment expert group composed of experts in various fields such as medicine, education and emergency management to collect data related to factor S_5. We collected data related to factor R_1 from the database of the epidemic prevention and control department. We asked that the head of epidemic control checks the item and fills out the online shared form questionnaire to collect data related to factors R_2, R_3, and R_4. We collected data related to factor R_5 from the duty record registration form.

3. Results

According to the case study, the percent of high risk level, medium risk level and low risk level of students' residence (P_1) were 6.71, 8.03, and 85.26%, respectively. The regional Health Commission thought that epidemic risk level of the school's location (P_2) is low risk level. The ways that students returned to school (P_3) were as follows: 8.6% by car, 64.3% by train (high-speed rail), 6.8% by taxi, and 20.3% by other means. The size of the university population (P_4) is 1,850. The number of migrants on and off campus (P_5) is 56, and it accounts for 10% of the total population. The high risk level, medium risk level and low risk level of express carrier infection (P_6) were 1.25, 2.91, and 95.84%, respectively. The level of students' knowledge of the epidemic (S_1) is 3.4. The percent of the very strong level, strong level, general level and weak level of students' awareness of risk (S_2) were 9.16, 81.67, 5.92, and 3.25%, respectively. Cleaned up campus health dead spots, disinfected public places, equipped with disinfectants and hand sanitizer were finished for the situation of campus environment renovation (S_3). The percent of the propaganda of anti-epidemic culture (S_4) on the school's official website, official account, Douyin, and other platforms is 75.68%. The expert group thought that the personnel control measures (S_5) were slightly defective. The epidemic prevention and control department asked the periodic nucleic acid testing (S_6) is three times a week. The expert group thought that the distance between the students is ∼1.2 m for campus space management (S_7). The satisfaction of teachers and students with online teaching based on the application

of digital teaching technology (S_8) is 91.28%. The percent of the student nucleic acid testing system registration is 88.43% for the application of digital epidemic prevention technology (S_9). The number of the health monitoring days per student in the data system (R_1) is 14 days. The expert group thought that the university conducted emergency drills for evaluation and made improvements based on evaluation comments (R_2). The expert group thought that the preparation of emergency plans (R_3) was comparatively perfect. There are two parties cooperating between universities and other subjects (R_4). Quarantine settings (R_5) were adequate quarantine areas and staff on full-day duty. Detected factors categories or values by subsystems (Pressure, state and response), with corresponding risks and weights, are reported in Tables 3–5.

After calculation, $H = 1.880$ and $V = 1.666$ were obtained, indicating that the comprehensive risk of the school's resumption was low. Therefore, risk prevention and control measures needed to be improved accordingly.

4. Discussions

The COVID-19 pandemic highlighted the need to multiply our efforts in epidemic prevention and control to protect public health. Since young people (often asymptomatic) are important spreader of COVID-19 (1), colleges and universities need to assess the risks involved in resuming school and make evidence-based decisions. To improve the risk index, we considered more factors compared to previous studies and described the influencing mechanisms between them through the PSR model. Finally, the disaster management perspective provided a clear picture highlighting the scale of university response. In effect, while the system vulnerabilities show where the countermeasures can be applied to be effective, the PSR model describes their impact.

4.1. Effective response suggested from the model

Based on our evaluation results, our study judged the risks involved in resuming school and put forward specific countermeasures for reducing the risk levels. This not only protects public health security but also has some practical implications for improving the evaluation and rational decision-making abilities of all parties. Students should apply to return to school in the system given by the school according to their requirements. They should fill in the date of their return and their mode of transportation, such as bus number and other relevant information, and they can only return after the school has provided their approval. Students are required to have health and travel codes, a 14-day health monitoring information form, and nucleic acid test proof within 48 h to enter the campus. Students are required to sign the Student Commitment to Return to School and strictly comply with the requirements of returning to campus. After entering the campus, the school will immediately disinfect students' luggage, bags, and other items, and conduct nucleic acid tests at designated locations. The school told students not to walk around the campus without special reasons and to narrow their scope of activities as far as possible. Students are not allowed to return to school without verification. Schools can hold lectures on epidemic prevention and control knowledge and relevant laws and regulations regularly, both online and offline. At the same time, information about epidemic prevention and control and national prevention and control policies should be posted on the school's publicity board and dormitory bulletin board. Because online teaching is adopted during closed-loop management, schools should actively take measures. Schools should strengthen the awareness of students and teachers regarding digital teaching technology and the use of teaching software so that they can correctly and skillfully use Dingding, Rain Classroom, Tencent Conference, and other platforms for teaching, ensuring the smooth progress of courses. Teachers should change their management mode from offline to online in a timely fashion. They should also make full use of the functions of each lecturing platform, such as check-in, links, video, and submitting homework, to innovate their teaching and enhance the effectiveness of their student management to ensure students continue with an appropriate level of learning engagement. However, the school should coordinate and improve the student management network platform system, collect all students' personal and facial information, and form a complete data management database. By doing so, the school can improve the accuracy of management, reduce the workload, and achieve high levels of management efficiency. The school should carry out refined prevention and control work in strict accordance with the relevant regulations of the national and provincial CDC, insist on regular nucleic acid testing and health reports, and strictly isolate migrants. The organization's institutional framework should also be optimized to clarify the responsibilities for epidemic prevention and control. Schools should ensure basic medical security and that they have adequate supplies, and implement real-name registration applications. School leaders and related management personnel should not only perform their respective duties but also give responsibilities to student party members. The school should organize students to be on duty every day in designated places, such as restaurants and libraries, to supervise students' daily behaviors so that disease prevention procedures can be fully implemented among students.

4.2. Response suggested to the case study

The results of the case study show that the university's comprehensive risk of resuming classes is low risk, and the

university can allow students to return to classes in terms of epidemic prevention and control. In terms of catastrophes, the university and most of its students were at a low risk level. The size of returning students and delivery of express have little impact on the epidemic prevention and control of the university. The transportation of students back to school is the most critical factor in the disaster. This requires the establishment of rules for returning students to school. In terms of vulnerability, campus epidemic prevention equipment, epidemic prevention culture publicity, health testing, epidemic prevention drills and quarantine Settings play an active role in prevention and control. However, risk awareness among students and digital epidemic prevention are hindering epidemic control. This requires increasing students' awareness of risks and using big data intelligence to enhance epidemic prevention. Other aspects should also be effectively addressed. In general, the university should further improve and optimize the transportation mode of students returning to campus, students' risk awareness and digital epidemic prevention to enhance the epidemic prevention and control effect. Based on the above analysis, the following measures are proposed. (1) Improve students' return to school information statistics. According to the arrival time of students by plane, train (high-speed railway) and other public transportation, school buses and special buses can be arranged at the airport, high-speed railway station and other transportation stations to reduce the risk of infection on the return trip. (2) The school vigorously publicized how individuals could contribute to epidemic prevention and control, provide role models, and create a cultural atmosphere for epidemic prevention and control on campus. Schools can conduct publicity through online platforms, shoot high-quality and positive short videos on epidemic prevention and control, and regularly release and update these materials on Douyin, Kuaishou, and other platforms to expand the scope of publicity and influence. In this way, students can improve their epidemic and risk awareness and regulate their behavior in strict accordance with institutional requirements to deal with the current severe situation with the correct attitude. (3) Schools should establish an epidemic prevention command platform and control center, and formulate a 24-h duty system for epidemic prevention and control. Additionally, on-duty staff should carry out training and education to ensure timely responses to all kinds of emergencies.

4.3. Limitations and strengths of the study

Our study had some limitations, which deserve further study and attention. First, the functions to evaluate the risk of factors derived from subjective evaluations although they are based only on the experience of the risk assessment expert group composed of experts in various fields such as medicine, education and emergency management. Second, our study only used one case study while comparisons could help to calibrate the measure tool. The research system should be expanded according to the varying situations of different universities and more indicators should be included in the evaluation process. However, our model (to the best of authors knowledge) is to date one of the most complete describing the complex interaction mechanism of factors that affects the university spread of the virus. In addition, artificial intelligence technology could be gradually introduced to find more factors and as support of factors weights assignment.

Data availability statement

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

Ethics statement

Ethics review and approval/written informed consent was not required as per local legislation and institutional requirements.

Author contributions

Conceptualization: SY, YW, and LM. Methodology: LM and YW. Software and writing—review and editing: LM and SY. Validation: YW and SY. Writing—original draft preparation: YW and TD. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This research was funded by the 11th batch of teaching research projects of Hebei Agricultural University in 2021, including research on interdisciplinary cultivation and reading practice teaching mode of agricultural college students (2021C-39) and the construction of ideological and political case base of organizational behavior (2021B-2-01), and the Chinese Association of Degree and Graduate Education, including research on the cultivation of the practical ability of graduate students majoring in agronomy (2020MSB37) and Hebei Agricultural University's first-class undergraduate course construction project Management System Engineering.

Acknowledgments

We are very grateful to Dr. Pierpaolo Ferrante for his warm help and inspiration. We also thank Dr. Xing Zeyu for his help.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.1037818/full#supplementary-material

References

1. Ferrante P. The first 2 years of COVID-19 in Italy: incidence, lethality, and health policies. *Front Public Health.* (2022) 10:986743. doi: 10.3389/fpubh.2022.986743

2. http://www.gov.cn/zhengce/zhengceku/2022-03/15/5679257/files/49854a49c7004f4ea9e622f3f2c568d8.pdf (accessed November 27, 2022).

3. Yin S, Zhang N, Xu J. Information fusion for future COVID-19 prevention: continuous mechanism of big data intelligent innovation for the emergency management of a public epidemic outbreak. *J Manag Anal.* (2021) 8:391–423. doi: 10.1080/23270012.2021.1945499

4. Zhang FR. Realization path of modernization of university emergency management from the perspective of PSR theory: based on the practice of COVID-19 outbreak resumption in Tianjin Normal University in 2020. *Gansu Educ Res.* (2021) 124–8.

5. Zhang F, Li S, Cui J, Tian Y, Li L, Yang J, et al. Emergency management in university students during coronavirus disease 2019 epidemics: West China urgent recommendation. *Chin J Evid Based Med.* (2020) 252—7.

6. http://www.moe.gov.cn/jyb_xwfb/xw_zt/moe_357/jyzt_2020n/2020_zt03/yw/202003/W020200312377387887254.pdf (accessed November 27, 2022).

7. Yang T. Construction of risk assessment system for respiratory infectious disease outbreaks in schools. *Chin Sch Health.* (2017) 38:1107–9.

8. Liu HL, Zhang GG. On the application of network security risk assessment in the overall framework of smart campus. *Netw Secur Technol Appl.* (2020) 97—9.

9. Ding H, Li JD. Risk assessment methods after school resumption in China in the context of COVID-19. *Henan Sci.* (2020) 1669—77.

10. Wang TN, Zhai Y, Wang K, Zhao RF, Han SP. Comprehensive risk assessment of university resumption under COVID-19. *China Saf Sci J.* (2021) 143–8.

11. Bai X, Tang J. Ecological security assessment of Tianjin by PSR model. *Proc Environ Sci.* (2010) 2:881–7. doi: 10.1016/j.proenv.2010.10.099

12. Zhou CX, Zhang T, Teng YZ, Liu JR. COVID-19 prediction based on BPNN and SEIR analysis of university resumption. *Chin J Soc Med.* (2020) 581—5.

13. Wang ZJ. Study on epidemic prevention and control countermeasures of university reopening. *Mass Standard.* (2020) 94–6.

14. Cui TF. Thoughts on ensuring normal school resumption after COVID-19 in Taiyuan city. *J Appl Med Technol.* (2020) 1693—4.

15. Yin S, Zhang N, Ullah K, Gao S. Enhancing digital innovation for the sustainable transformation of manufacturing industry: a pressure-state-response system framework to perceptions of digital green innovation and its performance for green and intelligent manufacturing. *Systems.* (2022) 10:72. doi: 10.3390/systems10030072

16. Li Y, Wei SH. Epidemic prevention and control strategies in colleges and universities. *Sci Technol Ind.* (2022) 109–13.

17. Dong YH. Campus management strategies in the context of COVID-19. *Jiangsu Sci Technol Inf.* (2020) 34—6.

18. Chen J. Epidemic prevention and control measures in communities and schools. In: *Chinese People's Political Consultative Conference.* (2022) 2022:5–6.

19. Jiang ZY. On the prevention and control measures of university libraries under the COVID-19 epidemic. *Inner Mongolia Sci Technol Econ.* (2021) 132—3.

20. Gao YQ. Research on security measures of colleges and universities in the context of COVID-19 normalized prevention and control. *Mod Occup Saf.* (2021) 71–3.

21. Wu H, Zhou SC, Zhang Q, Li XW, Hu JH. Prevention and control of COVID-19 among university graduate students and countermeasures. *Chin J Med Eng.* (2020) 30—3.

22. Hu XQ. Safeguard measures and effectiveness analysis of online teaching in colleges and universities during COVID-19. *Chin J Agric Educ.* (2020) 50—4.

23. Chen M, Guo YJ, Yu ZM. Improved type-order relation analysis and its application. *J Syst Manag.* (2011) 20:352–5.

24. Zhao XB, Wang Y. Flight quality risk measurement method based on skewness distribution. *China Saf Sci J.* (2019) 29:160–6.

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 13 January 2023
DOI 10.3389/fpubh.2022.1050673

OPEN ACCESS

EDITED BY
Pierpaolo Ferrante,
National Institute for Insurance Against Accidents at Work (INAIL), Italy

REVIEWED BY
Alfredo Maria Gravagnuolo,
Vatic Health Ltd., United Kingdom
Parnian Shobeiri,
Tehran University of Medical Sciences, Iran

*CORRESPONDENCE
Rosa María Gutiérrez-Ríos
✉ rosa.gutierrez@ibt.unam.mx

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Infectious Diseases: Epidemiology and Prevention, a section of the journal Frontiers in Public Health

RECEIVED 22 September 2022
ACCEPTED 28 December 2022
PUBLISHED 13 January 2023

Two-year follow-up of the COVID-19 pandemic in Mexico

Antonio Loza[1], Rosa María Wong-Chew[2†], María-Eugenia Jiménez-Corona[3†], Selene Zárate[4†], Susana López[1†], Ricardo Ciria[5], Diego Palomares[5], Rodrigo García-López[1], Pavel Iša[1], Blanca Taboada[1], Mauricio Rosales[1], Celia Boukadida[6], Alfredo Herrera-Estrella[7], Nelly Selem Mojica[8], Xaira Rivera-Gutierrez[1], José Esteba Muñoz-Medina[9], Angel Gustavo Salas-Lais[10], Alejandro Sanchez-Flores[11], Joel Armando Vazquez-Perez[12], Carlos F. Arias[1] and Rosa María Gutiérrez-Ríos[5]*

[1]Departamento de Genética del Desarrollo y Fisiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico, [2]Facultad de Medicina, Laboratorio de Investigación en Enfermedades Infecciosas, División de Investigación, Universidad Nacional Autónoma de Mexico, Ciudad de México, Mexico, [3]Departamento de Epidemiología, Instituto Nacional de Cardiología Ignacio Chávez, Ciudad de México, Mexico, [4]Posgrado en Ciencias Genómicas, Universidad Autónoma de la Ciudad de México, Ciudad de México, Mexico, [5]Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico, [6]Centro de Investigación en Enfermedades Infecciosas, Instituto Nacional de Enfermedades Respiratorias Ismael Cosío Villegas, Ciudad de México, Mexico, [7]Centro de Investigación y de Estudios Avanzados del IPN, Laboratorio Nacional de Genómica para la Biodiversidad-Unidad de Genómica Avanzada, Irapuato, Guanajuato, Mexico, [8]Centro de Ciencias Matemáticas, Universidad Nacional Autónoma de México, Morelia, Michoacan, Mexico, [9]Coordinación de Calidad de Insumos y Laboratorios Especializados, Instituto Mexicano del Seguro Social, Ciudad de México, Mexico, [10]Laboratorio Central de Epidemiología, Instituto Mexicano del Seguro Social, Ciudad de México, Mexico, [11]Unidad Universitaria de Secuenciación Masiva y Bioinformática, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico, [12]Instituto Nacional de Enfermedades Respiratorias Ismael Cosío Villegas, Ciudad de México, Mexico

Background: After the initial outbreak in China (December 2019), the World Health Organization declared COVID-19 a pandemic on March 11th, 2020. This paper aims to describe the first 2 years of the pandemic in Mexico.

Design and methods: This is a population-based longitudinal study. We analyzed data from the national COVID-19 registry to describe the evolution of the pandemic in terms of the number of confirmed cases, hospitalizations, deaths and reported symptoms in relation to health policies and circulating variants. We also carried out logistic regression to investigate the major risk factors for disease severity.

Results: From March 2020 to March 2022, the coronavirus disease 2019 (COVID-19) pandemic in Mexico underwent four epidemic waves. Out of 5,702,143 confirmed cases, 680,063 were hospitalized (11.9%), and 324,436 (5.7%) died. Even if there was no difference in susceptibility by gender, males had a higher risk of death (CFP: 7.3 vs. 4.2%) and hospital admission risk (HP: 14.4 vs. 9.5%). Severity increased with age. With respect to younger ages (0–17 years), the 60+ years or older group reached adjusted odds ratios of 9.63 in the case of

admission and 53.05 (95% CI: 27.94–118.62) in the case of death. The presence of any comorbidity more than doubled the odds ratio, with hypertension-diabetes as the riskiest combination. While the wave peaks increased over time, the odds ratios for developing severe disease (waves 2, 3, and 4 to wave 1) decreased to 0.15 (95% CI: 0.12–0.18) in the fourth wave.

Conclusion: The health policy promoted by the Mexican government decreased hospitalizations and deaths, particularly among older adults with the highest risk of admission and death. Comorbidities augment the risk of developing severe illness, which is shown to rise by double in the Mexican population, particularly for those reported with hypertension-diabetes. Factors such as the decrease in the severity of the SARS-CoV2 variants, changes in symptomatology, and advances in the management of patients, vaccination, and treatments influenced the decrease in mortality and hospitalizations.

KEYWORDS

COVID-19, variants, comorbidities, symptoms, logistic-regression, case-fatality-proportion

Introduction

In the last 2 months of 2019, cases of a novel severe pneumonia of unknown etiology were initially detected in the city of Wuhan, China. Using molecular biology techniques and genomic sequencing, its etiologic agent was characterized and classified as a new virus in the *Betacoronavirus* genus of the *Coronaviridae* family, phylogenetically closely related to severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV). This new virus was designated severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (1), and the associated disease was named coronavirus disease 2019 (COVID-19). After the initial outbreak in China and its subsequent epidemiological spread, the World Health Organization (WHO) declared COVID-19 a Public Health Emergency of International Concern on January 30th, 2020, and a pandemic on March 11th, 2020[1] Before scaling to a global pandemic, the fatality rate in China ranged between 2 and 3.7% (2), with a greater impact observed among older adults. This rate pattern was reproduced around the world, where older adults were the most affected group (3). Additionally, reports indicated that a set of common symptoms associated with COVID-19, including fever, cough, dyspnea, sputum production, headache, myalgia, and fatigue, have been reported worldwide (4). Symptoms with a lower prevalence in the population, including diarrhea, hemoptysis and difficulty breathing (5), were also reported. Nevertheless, symptoms have varied, and anosmia and dysgeusia were also acknowledged as potential clinical markers of the disease. Different SARS-CoV-2 variants have emerged throughout the pandemic due to the natural accumulation of mutations in the viral genome that have circulated worldwide. To monitor viral evolution, the WHO encourages genomic analysis of virus samples. Relevant variants associated with a risk of global impact ranging from possible to alarming are classified as under monitoring (VUM), of interest (VOI), and of concern (VOC)[1]. Physicians treating COVID-19 patients have reported changes in the symptomatology associated with specific variants detected in the studied patients. For example, patients with the Omicron variant, first detected in samples collected in South Africa on November 14th, 2021, showed symptoms that were more similar to those of a common cold, mostly without anosmia or dysgeusia (6). In Mexico, the first SARS-CoV-2-positive sample was reported on February 27th, 2020, from a patient returning from a trip to Italy. By March 4th, the number of cases had risen to 80, strongly suggesting the onset of community transmission of the virus (7). On March 30th, 2020, the Mexican government emitted a national epidemiological alert placing the general population under lockdown and suspending non-essential activities, allowing only those related to health, security, governance, services, and the economy. By March 19th, 2022, there were more than 5.7 million confirmed cases and 324 thousand deaths according to official epidemiological reports from the Mexican Health Ministry (SSA for its acronym in

Abbreviations: CCs, Confirmed cases; RT–PCR, Polymerase Chain Reaction test; CFP, Case fatality Proportion; HP, Hospital Admission Proportion; AR, Admission Risk; OR, Odd Ratio; SSA, Mexican Federal Health Ministry; DGE, General Directorate of Epidemiology; COVID-19, Coronavirus disease 2019; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; MERS-CoV, Middle East respiratory syndrome coronavirus; SISVER, Respiratory Diseases Surveillance System.

1 https://www.who.int/en/news/item/27-04-2020-who-timeline---covid-19

Spanish, Secretaria de Salud) that monitored cases through the Respiratory Diseases Surveillance System (SISVER) and the General Directorate of Epidemiology (DGE). The SISVER database includes clinical and epidemiological information that allows the tracking of the pandemic[2] Data available up to March 19th, 2022, show that the country has experienced several epidemiological surges (peaks or epidemiological waves) in the number of confirmed cases (CCs) of SARS-CoV-2 infection. During late 2020 and early 2021, Mexico registered a significant increase in the number of CCs, mainly driven by the B.1.1.519 variant (8), which was classified as a variant under monitoring (VUM) by the WHO in 2021. The vaccination campaign began for healthcare personnel and the over-60 population in the same period. Later, by June 2021, the pandemic was dominated by the Delta variant of concern (VOC) (9), while three other vaccines with different levels of effectiveness, as seen in Supplementary Table 1, were added to the vaccine campaign. Since December 2021, the Omicron variant and its sub-variants have been circulating in Mexico. This study aimed to detail the epidemiological evolution of the COVID-19 pandemic in Mexico during the first 2 years (March 2020–2022), the health policies adopted by the national government, and the circulating virus variants.

Methods

Settings

During the week of March 30th, 2020, the DGE declared a public health emergency and stated that the country had entered a community transmission stage. The strategy followed by the Mexican government for the epidemiological surveillance of COVID-19 is represented in Figure 1. The first step in the diagnosis began when a patient attended a health care unit after being in contact with an infected individual or showing symptoms. After filling out the admission form and questioning, the health personnel determined whether the patient should be tested. At the beginning of the pandemic, the patients were tested for SARS-CoV-2 only in hospitals. Then, the Mexican government published a standard that makes COVID-19 tests available to clinical laboratories and drugstores. Until mid-November, RT–PCR was the only test available, and antigen testing was authorized by Mexican health authorities. Six vaccines were used during the vaccine campaign in this period[3], and the main characteristics and dates of approval are summarized in Supplementary Table 1.

Study population and design

This is a population-based longitudinal study.

Participants

All suspected COVID-19 cases recorded in the SISVER database.

Outcomes

Number of suspected cases (SCs), number of confirmed cases (CCs), number of CCs who died (deaths), number of CCs who were hospitalized (hospital admissions), case fatality proportion (CFP), and hospital admission proportion (HP).

Independent variable

Sex (men, women), age (0–17 yrs., 18–29 yrs., 30–39 yrs., 40–49 yrs., 50–59 yrs., 60+ yrs.), patient comorbidities (hypertension, obesity, diabetes, asthma, heart disease, renal insufficiency, COPD, immunosuppression, HIV/AIDS), patient symptoms, virus variant (lineages recorded in GSAID), and monthly percentages of vaccinated people.

Data source/measurement

The data provided by the SSA (through the DGE) contain suspected or confirmed cases, which include ambulatory, hospitalized, and deceased patients with demographic variables, self-reported comorbidities, and the main symptoms. **SCs** are patients seeking medical care as suspects (with symptoms or after contact with a CC). **CCs** are individuals with a quantitative reverse transcription polymerase chain reaction test (RT–PCR) positive for SARS-CoV-2, positive antigen tests or a positive result ruled by epidemiologic association (confirmed cases by epidemiologic association). **Confirmed cases by epidemiologic association** are SCs who have been in close contact (living within a distance of less than 1 meter for 15 continuous or cumulative minutes) with a laboratory-confirmed case by RT–PCR or rapid antigen test for SARS-CoV-2, from 2 to 14 days before the onset of symptoms and that the confirmed case to which it is associated, is registered on the SISVER platform or in the Online Notification System for Epidemiological Surveillance (SINOLAVE). **Deaths** are the CCs that held a death certificate. Hospital admissions are the CCs that were hospitalized. **CFP** is the fraction of deaths among the CCs. **HP** is the fraction of hospital admissions among the CCs. The operational definitions of outcomes were taken from the Mexican standard for epidemiological surveillance[4] We chose the age groups following the vaccination strategy implemented by the Mexican government. The reported comorbidities were obtained through the suspected case study form completed during admission or health care visits. Patient symptoms were recorded by the health care personnel. For each suspected case

2 https://coronavirus.gob.mx/informacion-accesible/

3 https://covid19.trackvaccines.org/country/mexico/

4 https://coronavirus.gob.mx/wp-content/uploads/2022/01/2022.01.12-Lineamiento_VE_ERV_DGE.pdf

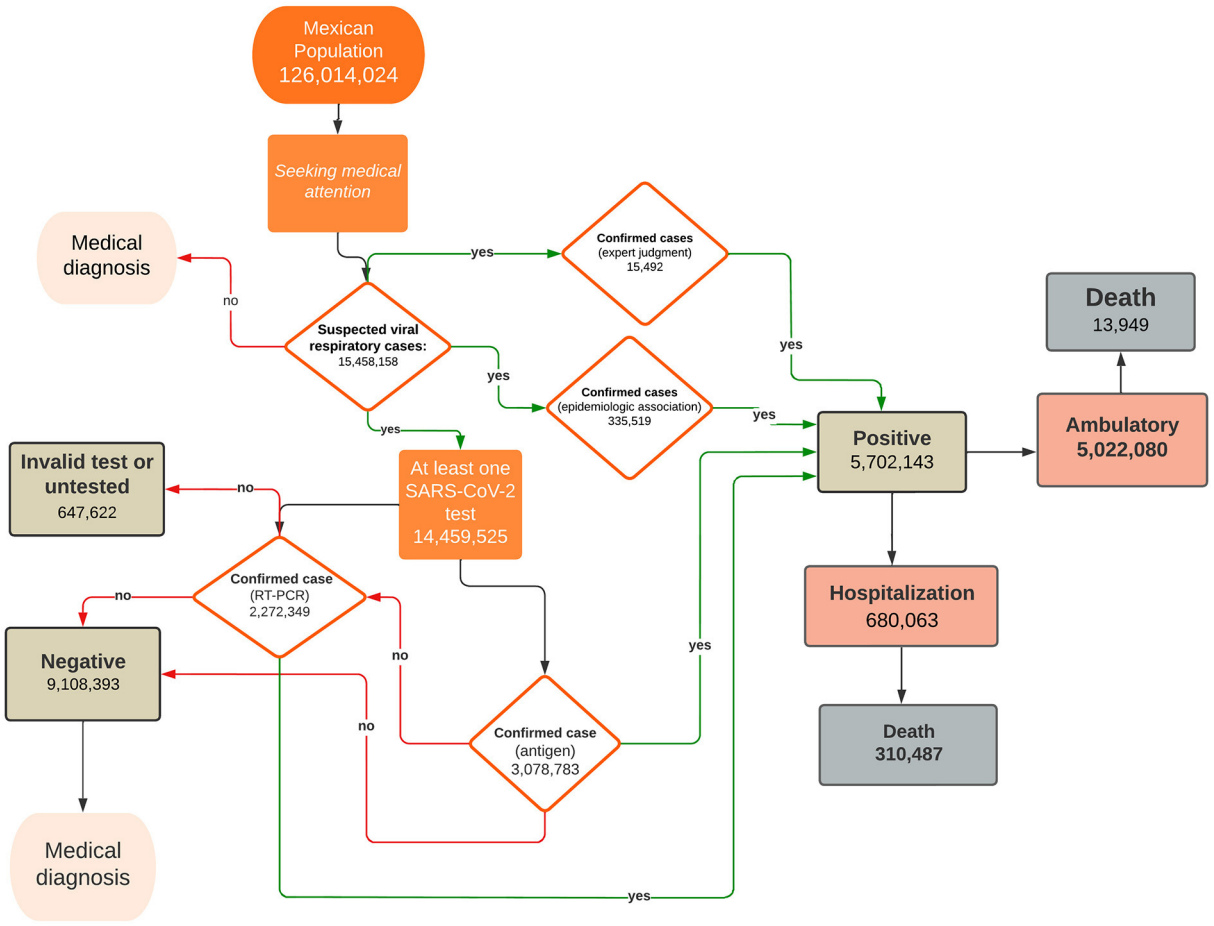


FIGURE 1
Follow-up of suspected COVID-19 cases reported in the SISVER database. The diagram shows suspected cases of viral respiratory disease and their evolution path. A fraction of the patients with a negative antigen test were tested by RT–PCR.

tested, one or more tests can be conducted, but the data set in our study reports only the last result. Sequenced SARS-CoV-2 genomes from Mexico were uploaded to the GISAID database[5] that had assigned lineage and date of complete sample collection ($n = 47{,}572$). Those sequences represent less than 1% of the CCs. Monthly percentages of people who received one dose of a vaccine and those who were fully vaccinated were downloaded from "Our World in Data"[6]

Statistical analysis

The period analyzed in this work comprises epidemiological week 14 of 2020 (beginning on March 29th, 2020) up to epidemiological week 11 of 2022 (concluding on March 19th, 2022). Data were grouped according to the date of symptom onset. The distributions of the number of CCs, deaths and hospital admissions were analyzed by epidemiological period, sex, and age. Using the last result for every patient tested, we assessed a lower bound for the weekly number of tests (Supplementary Figure 2). The epidemic peaks were determined considering the changes in CC numbers in a three-week moving average of the weekly growth factor G_n, where n means the nth week, which is calculated as the difference in natural logarithms (ln) of new cases accumulated in two subsequent weeks:

$$G_n = \ln\left(N_I\ (t_n)\right) - \ln\left(N_I\ \left(t_{n-1}\right)\right)$$

where $N_I\ (t_n)$are the new cases reported during week nth. This approach was chosen because, in the early stages of any epidemic, the number of infected patients grows exponentially at a given rate of G (10); this implies that the number of new infections in a time interval of length t is approximately

5 https://gisaid.org/hcov19-variants/

6 https://ourworldindata.org/grapher/covid-vaccination-doses-per-capita?country~MEX

expressed by I(t) α exp(Gt). Nevertheless, the weekly growth factor G_n is also helpful in obtaining information on contagion dynamics in every step of the pandemic. To characterize the waves, we performed descriptive analysis using simple frequencies and percentages of study variables. The CFP and HP were estimated as the average of 100 subsamples of size 15,000 taken from the original data set. After applying the Shapiro–Wilks test, we assumed the data's normality and calculated the 95% CI. To show the dynamics of the SARS-CoV-2 lineages that circulated in our country, a pile density curve was built. To show how symptoms have changed over time and how frequently they were among patients, we carried out cluster analysis. For the most frequent comorbidities among the Mexican population (diabetes, hypertension, and obesity), we also explored the combined effects of each pair of those comorbidities. Finally, we used multivariate logistic regression models with death and hospital admission as outcomes and the epidemic wave (1, 2, 3, 4), sex, age and the presence of comorbidities (yes, no) as risk factors. All analyses were performed with R v.4.1 statistical software[7] We used ggplot2 v.3.3. to build the pile density curve and the *pheatmap* package[8] with default options and the *complete* option to group symptoms throughout the study period.

Results

At the end of this study (March 19th, 2022), the national COVID-19 registry included a total of 15,458,158 suspected cases, out of which 5,702,143 were CCs, while 9,108,393 were not. The remaining 647,622 had no reported result because they were not tested or because the result was considered invalid (for example, due to poor sampling or poor handling of the test) (Figure 1). The overall fraction of CCs among the SCs (r_c) is equal to 0.36, but it strongly varied over time and reached its maximum (0.67) in the last weeks of the study period. Supplementary Figure 1 shows that the peaks of r_c and epidemic waves approximately coincide. The curves of CCs, hospitalizations and deaths showed four peaks, but while those of CCs tended to increase, the others tended to decrease. The vaccination campaign started at the end of 2020, and in 1 year, 63.6% of the population had at least one dose (Figure 2A). Several virus variants became prevalent, with each time the latest replacing the previous one at a faster rate (Figure 2B). Of all CCs, 680,063 were hospitalized (11.9%), and 324,436 (5.7%) died. Even if the gender ratio of women to men was 1.08, males had a higher risk of death (CFP: 7.3 vs. 4.2%) and hospital admission risk (HP: 14.4 vs. 9.5%). Age was most strongly associated with the risk of death and admission. For the 60+ age group, CFP (26.8%) and HP (43.1%) were the highest; these values gradually decreased for the rest of the age groups. Remarkably, the HP curve by age is "J" shaped, with the 0–17 years group showing a higher HP (4.0%) than the next group.

Hypertension (12.7%), obesity (10.5%), and diabetes (9.5%) were the most prevalent comorbidities. Other comorbidities, such as heart disease, chronic obstructive pulmonary disease (COPD), renal insufficiency, and different immunosuppressive conditions, contributed to low percentages (<10%) among CCs. The percentage of deaths among people with diabetes was 21.9%, and that among people with hypertension was 19.8%, which was higher than the global percentage (14.5%). We also observed that the CFP of less prevalent comorbidities, such as renal insufficiency (38.1%), COPD (32.8%), and immunosuppression (21.6%), indicates an augmented risk of death (Table 1). In Figure 3, cluster (a) shows symptoms with a prevalence over 50% (cough, headache, fever, odynophagia, myalgias, and arthralgias); cluster (b) shows symptoms with a prevalence between 30 and 50% among people with CCs (rhinorrhea, chills, and sudden onset symptoms); and cluster (c) shows symptoms with a prevalence lower than 30% among people with CCs (vomiting, cyanosis, polypnea, abdominal pain, conjunctivitis, shortness of breath, chest pain, anosmia, dysgeusia, irritability, and diarrhea).

Determining the waves

We choose the date when the DGE declared the public health emergency (March 30th, 2020) as the starting day of the first wave. From that point, the weekly growth rate (G_n) decreased from values that were over 50% to negative values (when the trend was inverted and began a brief period of decline in the number of CCs). After a period with $G_n \approx 0$, there was a sudden increase at the beginning of week 40 (mid-September 2020) that marked the start of the second wave. The end of this second wave was followed by a five-week-long decline in CCs ($0\% > G_n > -12\%$). Afterward, as of epidemiological week 21 in 2021 (beginning on May 23rd, 2021), the beginning of the third epidemic wave was determined since G_n changed its sign. After a period of a moderate decrease at the end of the third wave ($0 > G_n > -7\%$), a substantial increase was observed ($G_n \approx 100$ %) in week 51 of 2021 (beginning on December 19th, 2021), becoming the fourth epidemiological wave. The $G_n$value showed substantial variation throughout the examined waves. The relatively high values at the beginning of the pandemic constituted a transient phenomenon. Afterward, the maximum observed G_n value during the first and second epidemic waves was approximately 20%, rising from more than 30% in W3 to almost 100% during the fourth wave (Figure 4).

7 https://cran.r-project.org/

8 https://CRAN.R-project.org/package=pheatmap

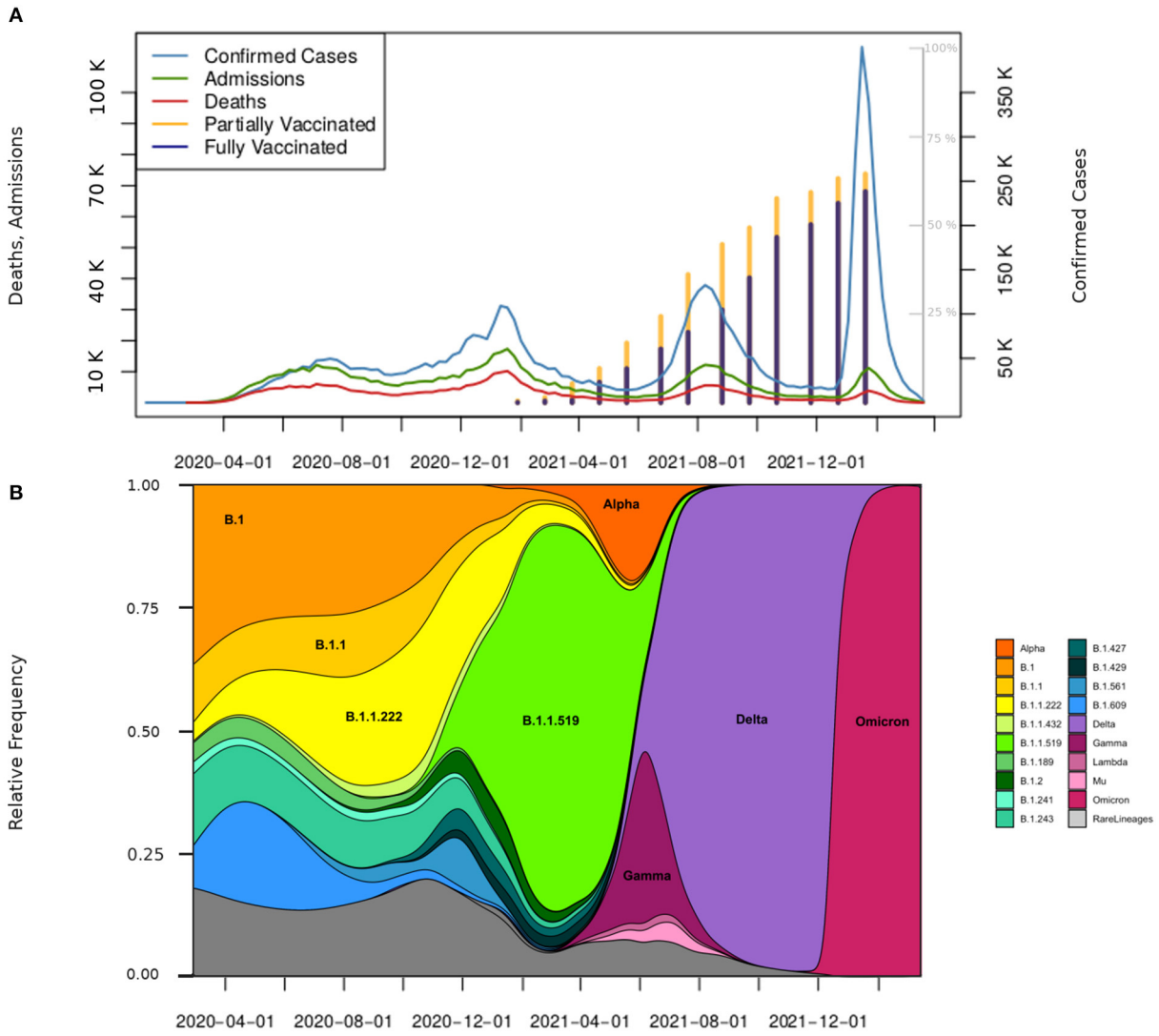


FIGURE 2
Confirmed cases, hospitalizations, deaths, vaccinations, and the presence of variants in Mexico. **(A)** Weekly distribution of CCs (scale on the right), hospitalizations, and deaths (scale on the left). The stacked bar plot in this figure presents the share of people partly (yellow bar) and fully vaccinated (navy bar). The percentage scale is displayed in gray on the inner right area of the plot. Bars are placed at the midpoint of the respective month beginning in January 2021. **(B)** Distribution of variants in the study period. The main variants that circulated in Mexico in the four evaluated periods are shown.

Wave 1 (W1)

The first wave started on March 29th, 2020 and ended on September 26th, 2020. During this period, 1,670,308 patients were tested at least once, and 809,387 CCs were detected with a median age of 43.7 years. Among CCs, there were 100,228 deaths (CFP = 12.3%) and 203,992 hospitalizations (HP = 25.1%) (Table 2). Approximately 150,000 CCs occurred in each 10-year age group over 18 and approximately 25,000 among children and adolescents (0–17 years). Hospitalizations and deaths increased exponentially with age (Supplementary Figure 3). The most common symptoms were headache, fever, myalgia, arthralgia, general malaise, and odynophagia (Figure 3). None of the circulating variants dominated during this first wave, and the most common variants included B.1, B.1.1, and B.1.1.222, from these, the first and the third variants reached its maximum prevalence (23%) in August-September 2020 (Figure 2B). The main strategy established by the Mexican government during this first wave was called the Safe Distance National Campaign, which began on March 23rd and ended on May 30th, 2020 (Figure 4). After this measure, a remarkable decrease in weekly growth rate (G_n) from values that were over 50% to values close to 20% ($G_n \lesssim 23\%$) was observed during mid-May. This

TABLE 1 Characteristics of COVID-19 infections, deaths, and hospitalization.

Disease	Population[a]		CCs		Deaths		Hospitalizations	
	n	%	*n*	%	*n*	CFP (%)	*n*	HP (%)
Gender								
Male	61,473,390	(48.8)	2,734,533 (Median age 38.4 yrs.)	(48.0)	199,655 (Median age 63.7 yrs.)	(7.3)	395,354 (Median age 58.3 yrs.)	(14.5)
Female	64,540,634	(51.2)	2,967,610 (Median age 38.4 yrs.)	(52.0)	124,781 (Median age 65.0 yrs.)	(4.2)	284,709 (Median age 58.8 yrs.)	(9.6)
Age (years)								
0–17	38,521,344	(30.6)	369,277	(6.5)	1,262	(0.3)	14,822	(4.0)
18–29	24,729,112	(19.6)	1,395,260	(24.4)	5,418	(0.4)	35,030	(2.5)
30–39	18,441,103	(14.6)	1,295,787	(22.7)	15,767	(1.2)	63,795	(4.9)
40–49	16,445,999	(13.0)	1,106,120	(19.4)	37,089	(3.3)	104,002	(9.4)
50–59	12,733,490	(10.1)	793,746	(13.9)	66,048	(8.3)	142,813	(18.0)
60+	15,142,976	(12.0)	741,953	(13.0)	198,852	(26.8)	319,601	(43.1)
Comorbidity								
Hypertension	-	-	722,714	(12.7)	143,429	(19.8)	250,145	(34.6)
Diabetes	-	-	542,746	(9.5)	119,071	(21.9)	212,366	(39.1)
Obesity	-	-	599,034	(10.5)	67,261	(11.2)	132,498	(22.1)
Asthma	-	-	109,701	(1.9)	55,57	(5.1)	13,319	(12.1)
Heart disease	-	-	61,180	(1.1)	16,151	(26.4)	28,249	(46.1)
Renal insufficiency	-	-	60,275	(1.1)	22,986	(38.14)	37,618	(62.4)
COPD	-	-	42,917	(0.7)	14,084	(32.82)	23,377	(54.4)
Immuno-suppression	-	-	33,820	(0.6)	73,18	(21.64)	14,246	(42.1)
HIV/AIDS	-	-	16,594	(0.39)	1,506	(9.1)	3,297	(19.9)
Other	-	-	87,820	(1.5)	16,448	(18.7)	31,231	(35.6)
Overall*			1,403,710	(24.6)	203,104	(14.47)	380,477	(27.1)

Mexico March 2020-2022. [a] Population census of 2020. "-" Unknown. "*" Refers to all comorbidities.

G_n value remained positive and below 23% for 2 months until reaching its highest point in the 2nd week of July 2020. At this turning point ($G_n = 0$), the trend was inverted and began a brief period of decline in the number of CCs followed by a ($G_n \approx 0$) plateau.

Wave 2 (W2)

The second wave started in mid-September 2020 [when a new sustained increase in CCs was observed (Figure 2)] and ended in the 2nd week of April 2021. In this period, 4,302,882 patients were tested, and 1,538,110 CCs were detected with a median age of 41.9 years. Among CCs, there were 132,638 deaths (CFP = 8.7%) and 705,673 hospitalizations (HP = 16.4%). With respect to W1, the overall proportions of deaths and hospitalizations were reduced by one-third (Table 2), while in all age classes, the number of CCs, hospitalizations and deaths increased (Supplementary Figure 3). The symptoms clusters a and b showed a steep decrease in prevalence (Figure 3). The initial prevalent variant B.1.1.222 was gradually replaced by the B.1.1.519 variant, and neither has been considered a VOC. During November 2020, the government authorized the use of antigen tests as a method to confirm the infection (Supplementary Figure 2), and a month later, it started the vaccination campaign. After the peak of January 2021, the number of weekly CCs decreased drastically. The increase in CCs was mainly associated with the dissemination of the B.1.1.519 variant (Figure 2B).

Wave 3 (W3)

The third wave began as of epidemiological week 21 in 2021 (beginning on May 23rd, 2021) and ended on November 6th, 2021. During this period, 4,289,906 patients were tested, and 1,439,463 CCs were detected with a median age of 34.7 years. Among CCs, there were 61,155 deaths (CFP = 4.2%) and

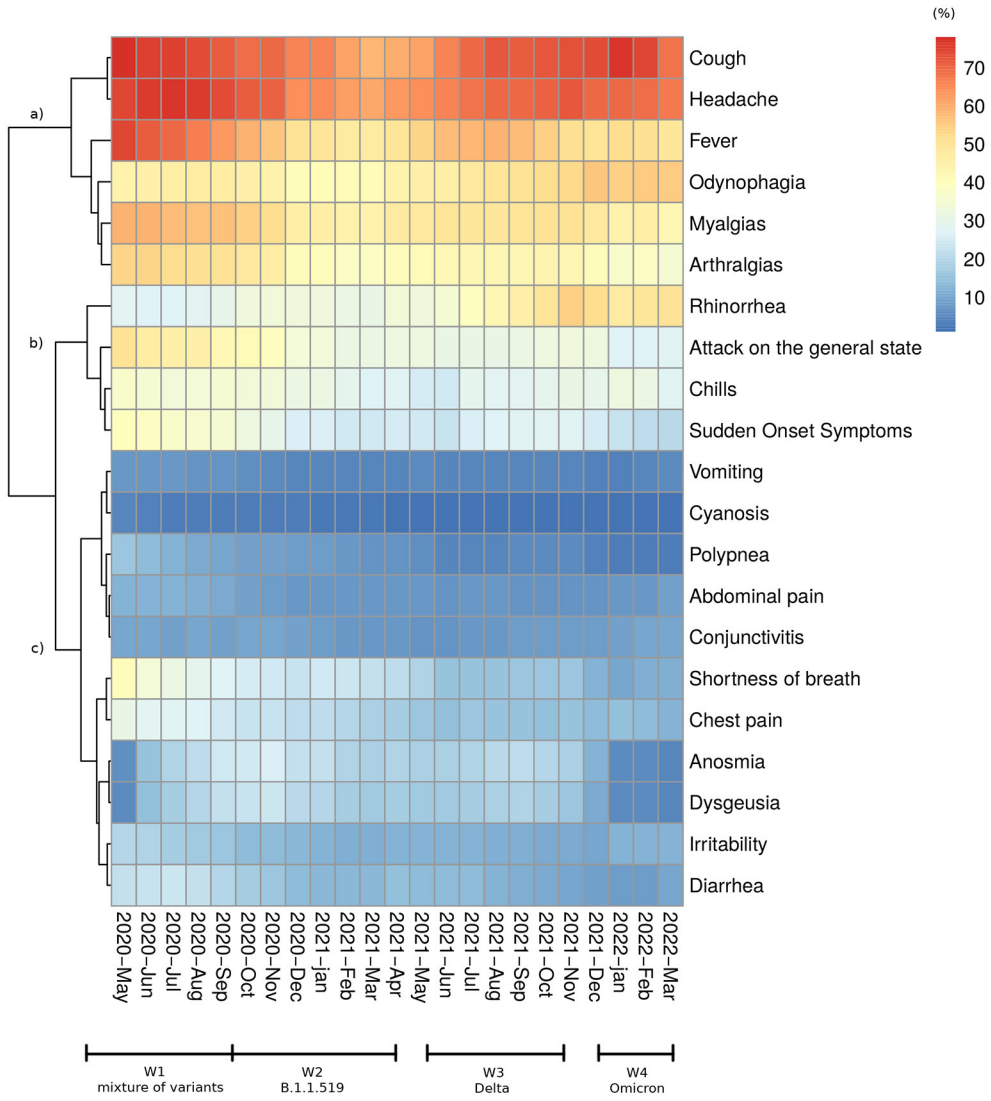


FIGURE 3
Symptom distribution. Heatmap and the cluster showing changes in the proportion of each recorded symptom among confirmed patients from May 2020 to March 2022.

141,067 hospitalizations (HP = 9.8%). With respect to W2, the proportions of deaths and hospitalizations halved with respect to the previous wave (Table 2). CCs increased in individuals under 40 years of age, were stable in the 40–49 year age class and decreased in people aged 50 years of age and older, while deaths and hospital admissions declined in all ages and above all among older adults (Supplementary Figure 3). The data show an increase in rhinorrhea and odynophagia, such as a new increase in fever, cough, headache, and a decrease in shortness of breath and chest pain. Although Alpha and Gamma variants initially replaced the B.1.1.519 lineage, the Delta variant (appeared in June 2021) quickly became dominant (87% prevalence in August 2021) and characterized this wave (Figure 2B). The vaccination campaign progressed with the inclusion of individuals between 18 and 29 years of age. The maximum of this wave was followed by a decrease in CCs that ended in week 44 of 2021.

Wave 4 (W4)

On December 19th, 2021, the fourth epidemiological wave started (Figure 4). As of March 19, 2022, 3,035,537 patients were tested, and 1,722,625 CCs were detected with a median age of 36.5 years. Among CCs, there were 20,659 deaths (CFP = 1.2%) and 58,569 hospitalizations (HP = 3.4%). With respect to W3, the proportions of deaths and hospitalizations were reduced by two-thirds (Table 2), and CCs increased in adult age classes (18–59 years), while they slightly decreased in youngest (0–17 years) and older adults (60+ years). Deaths

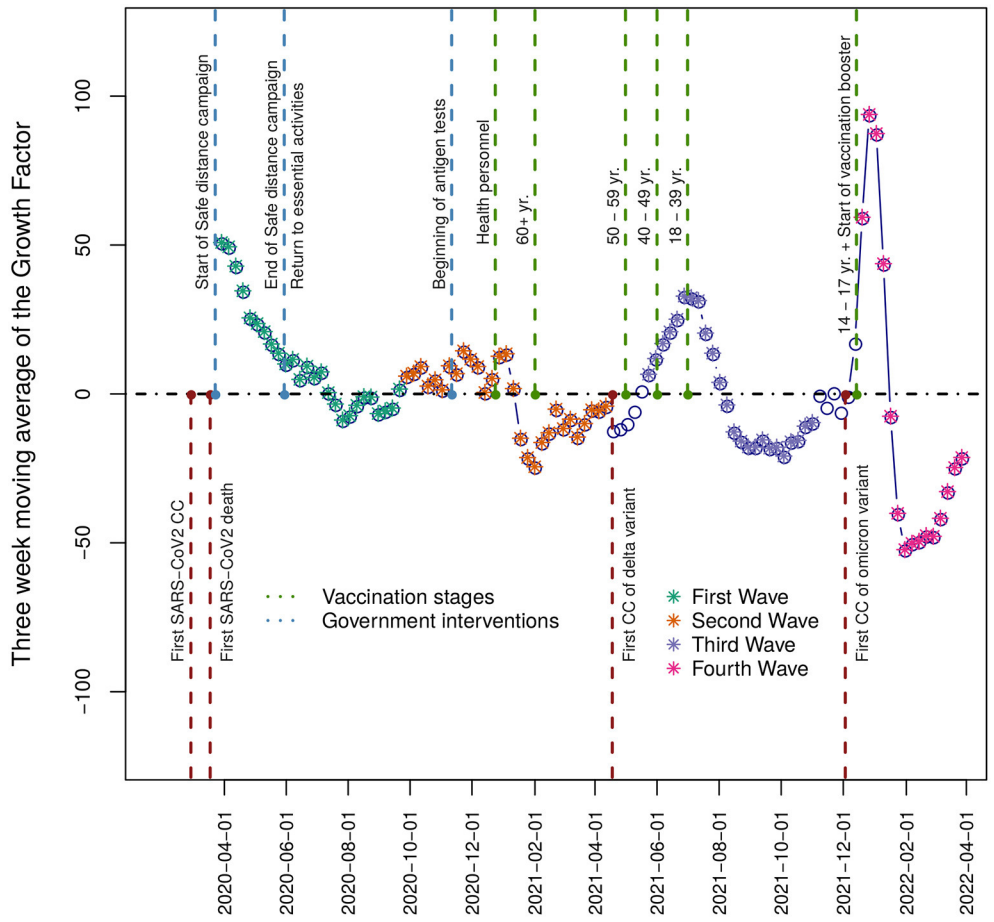


FIGURE 4
Growth Factor. The plot displays in the vertical axis the three-week moving average of the growth factor time series for the period under study—relevant events such as the progress in vaccination official strategy, variant detection, and public policy health interventions are shown. The filled points represent each wave, and the empty points represent the interwave periods.

and hospital admissions continued to decline in all age groups (Supplementary Figure 3). The prevalence of anosmia and dysgeusia decreased. Infections in this wave were driven by the Omicron variant BA.1, which replaced the Delta variant very quickly (Figure 2B). The peak in weekly CCs was reached between the second and third week of January 2022 and lasted up to the fourth week of February 2022, when the CCs showed a steep fall ($G_n < -50\%$) (Figure 4). The vaccination campaign included the first dose for individuals between 14 and 17 years and the booster shot for those aged 30 and over.

Multivariable analysis of hospital admissions and deaths in Mexico during the 4 waves

Taking as reference the age group between 0 and 17 years, the group of 60 years or older reached a maximum in both admission and death risk with adjusted odds ratios of 9.63 (95% CI: 7.22, 13.11) and of 53.05 (95% CI: 27.94, 118.62), respectively. In both cases, the ORs follow a descending pattern (Tables 3, 4). Interestingly, for the age group of 18 to 29 years, and despite not being significant in the cases of deaths, our results for admissions OR = 0.52 (CI 95%: 0.37, 0.73) and for deaths OR = 0.95 (CI 95%: 0.46, 2.25) imply a reduction in the admission and death risk concerning the reference group. We also observed a progressive reduction in the admission and death risk as the four waves elapsed, with the fourth wave displaying a stronger association with the decrease in admission and death risks. For this wave (W4), the fitted values for ORs in the admissions and deaths case were equal to 0.15 (95% CI: 0.13–0.18). Overall, men had higher odds of admissions (OR = 1.59; 95% CI: 1.44, 1.75) and deaths (OR = 1.78; 95% CI: 1.60, 1.97). Finally, the presence of comorbidities was associated with an increased admission (OR = 2.40; 95% CI: 2.17, 2.66) and death (OR=2.38; 95% CI: 2.14, 2.66) risk. For each combination of hypertension, diabetes and obesity, we

TABLE 2 Epidemiological waves.

Wave	Week starting	Week* ending	Tested** (Tested per week)	Confirmed cases	Hospital admissions	Deaths	Case fatality proportion (95% CI)	Hospital admission proportion (95% CI)
W1	14 March 29th, 2020	40 September 26th, 2020	1,670,308 (61,863)	809,387	203,992	100,228	12.3% (10.8–13.7)	25.1% (23.0–27.1)
W2	41 September 27th, 2020	15 April 17th, 2021	4,302,882 (148,375)	1,538,110	251,245	132,638	8.7% (7.8–9.5)	16.4% (15.3–17.6)
W3	21 May 23rd, 2021	44 November 6th, 2021	4,289,906 (178,746)	1,439,463	140,266	61,155	4.2% (3.6–4.9)	9.8% (9.1–10.6)
W4	51 December 19th, 2021	11 March 19th, 2022	3,035,537 (202,369)	1,722,625	60,021	20,659	1.2%(0.9–1.5)	3.4% (2.8–4.0)

The four epidemic waves were established using the Gn value. The table shows the epidemiological weeks covered by each wave, start and end dates, and the number of confirmed cases, hospitalizations, and deaths, excluding the interwave periods. The CFP and HP values are the averages of 100 subsamples of size 15,000 taken from the original data set. After applying the Shapiro–Wilks test, we assumed the data's normality and calculated the 95% CI * First week of each year: a) 2020: 2019-12-29; b) 2021: 2021-01-03; c) 2022-01-02. Weeks start on Sunday. ** Lower bound for the number of tests conducted.

TABLE 3 Hospital admissions risk.

	Odd ratio	95% CI	*P* value
Wave			
Wave 4	0.15	0.12 to 0.18	<2e-16
Wave 3	0.52	0.45 to 0.60	1.8E-13
Wave 2	0.72	0.64 to 0.81	1.0E-06
Wave 1	1.0	-	-
Sex			
Men	1.78	1.60 to 1.97	<2e-16
Women	1.0	-	-
Age			
> 59 yr.	53.05	27.94 to 118.62	<2e-16
50–59 yr.	15.35	8.05 to 34.43	6.6E-11
40–49 yr.	6.69	3.49 to 15.06	1.5E-06
30–39 yr.	2.67	1.37 to 6.08	0.0105
18–29 yr.	0.95	0.46 to 2.26	0.9004
0–17 yr.	1.0	-	-
Comorbidities			
Yes	2.38	2.14 to 2.66	<2e-16
No	1.0	-	-

Multivariable logistic regression results for risk of hospital admission among patients with coronavirus disease.

observed, throughout the waves, a downward trend in the percentage of COVID-19 patients with those comorbidities (Supplementary Figure 5). We can note a similar behavior in the case of deaths, except for the group of patients with both diabetes and hypertension, for which we observe an increased contribution to the total deaths in the fourth wave. We also note that the proportion of all these comorbidities in the CC increases by at least a factor of two in the total number of deaths for all the waves.

Discussion

From the beginning of the pandemic to March 29th, 2022, there were a total of 490,204,256 confirmed cases and 6,173,572 deaths around the world[9] The highest number of infections favored the appearance of new variants with some evolutionary advantage. The local emergence and dominance of SARS-CoV-2 variants as well as the health system responses modeled the pattern of the pandemic in the COVID-19 epidemiological profiles of countries (11). In Mexico, the first case of COVID-19 was recorded on February 27, 2020. For almost 1 month, the detected infections were

9 https://covid19.who.int/

10.3389/fpubh.2022.1050673

TABLE 4 Death risk.

	Odd ratio	95% CI	*P* value
Wave			
Wave 4	0.15	0.13 to 0.18	< 2e-16
Wave 3	0.46	0.40 to 0.53	< 2e-16
Wave 2	0.59	0.52 to 0.66	5.9E-13
Wave 1	1.0	-	-
Sex			
Men	1.59	1.44 to 1.75	1.9E-14
Women	1.0	-	-
Age			
>59 yr.	9.63	7.22 to 13.11	< 2e-16
50–59 yr.	3.18	2.37 to 4.35	6.0E-11
40–49 yr.	1.68	1.25 to 2.30	0.001275
30–39 yr.	0.93	0.68 to 1.28	0.630152
18–29 yr.	0.52	0.37 to 0.73	0.000205
0–17 yr.	1.0	-	-
Comorbidities			
Yes	2.40	2.17 to 2.66	< 2e-16
No	1.0	-	-

Multivariable logistic regression results for risk of death among patients with coronavirus disease.

all imported. The first local transmissions were reported on March 23rd, 2020, 1 week later, when the government declared a public health emergency (12); since then, and until March 2022, four epidemic waves have occurred, in contrast to Italy, where five waves and VOCs were reported (11).

Virus spread and evolution

As in the rest of the world, at the onset of the pandemic, a patchwork of virus variants circulated in Mexico (7). In the second wave, B.1.1.519 was the dominant variant. Rodriguez-Maldonado et al. (8) reported a sequence mutation at position T478K in the S protein (8) that may be involved in immune evasion and transmission advantage over the previous circulating variants. At the end of W2, the Alpha variant appeared first and spread faster than B.1.1.519, followed by Gamma, which spread even faster in some areas of the country (13). Finally, the introduction of the Delta variant occurred during mid-June 2021 (9), which reached 87% prevalence in August 2021 during the third wave peak (Figure 2B). In the fourth wave and as of December 2021, the Omicron variant (BA. 1) pushed out the Delta variant and became the most prevalent in March 2022 (the end of this analysis), representing over 90% of the sequences obtained (Figure 2B). As reported in several studies, all the VOCs showed each time an increased transmissibility with respect to the previous one (14). Even if the distribution of CCs over time also depends on the health policies adopted at the national and local levels, such as on the behavior of the population, the dynamic of the CCs from W2 to W4 (associated with the prevalent variants) is consistent with the ever-greater spread capacity developed by the variants that have followed one another. The evolution of the virus also altered patients' manifested symptomatology. At the onset of the pandemic (4), the most frequent symptoms were "similar to that of an acute respiratory infection", such as headache, fever, myalgia, arthralgia, general malaise, and odynophagia (Figure 3). Since anosmia and dysgeusia were poorly associated with other coronaviruses, these symptoms were not considered for diagnosis in the surveillance of W1, thus hindering early detection and treatment. The progression of the pandemic and of cases of the B.1.1.519 variant showed a decrease in symptoms such as cough and headaches. In line with results from other studies (15), symptoms such as rhinorrhea and odynophagia were more prevalent with the Delta variant. Instead, cough, fever, myalgia, malaise, headache, body ache, and moderate to severe fatigue were more common with Omicron (W4), supporting the assumption that this variant infects mainly the upper respiratory tract (16). Our data also confirmed (Figure 3) that anosmia was less prevalent in Omicron infections (17) and indicate that diagnosis is a challenge to physicians as new variants emerge.

Health policies and health system response

The maximum value recorded in the first wave for the growth factor G_n was the lowest for the four waves. This behavior could be related to the Safe Distance National Campaign proposed by the Mexican government on March 23rd, 2020. The campaign included school lockdowns and reduced economic activities, retaining only essential services. However, the were a relative limited number of diagnostic tests, thus reducing the detection of cases. Furthermore, the results of the logistic regression show that this wave presents, globally, both the highest admission and death risk. This behavior is confirmed in a study showing the leading causes of excess mortality in Mexico during 2020–2021 (18) and suggests that the safe distance campaign was a useful measure to reduce the number of CCs but had less impact on the proportion of hospitalized and deceased patients. W2 showed an increase in CCs compared with W1 (Figure 2), partly due to the higher number of total infections and the improved detection of infections. On the one

hand, fewer restrictions on population mobility increased the contact rate. On the other hand, the introduction of rapid tests increased the total number of tests conducted daily and allowed for more comprehensive monitoring of the pandemic. As shown in Table 2, W2 is also characterized by the highest number of deaths and hospital admissions and by the highest ratio of deaths/admissions (>0.5). Several reasons may have contributed to this result, such as hospital saturation and high occupancy of intensive care units in public and private institutions, as well as an increase in in-house oxygen demand. In W3, Delta dominance caused an increase in infections among younger ages that were not yet (0–17 years) or were just (18–29 years) included in the vaccination campaign. In contrast, infections among individuals over 50 years of age (the first to complete the vaccine cycle) decreased. Even if several studies found that Delta was the most virulent VOC (19), there was a decrease in hospital admissions and deaths that can be explained by several factors. One of them is the progress in the vaccination campaign, which for the first time included individuals between 18 and 59 years. By the end of July 2021, 16 and 20% of the total population were partially and fully vaccinated, respectively. In this period, other vaccines were introduced in the vaccine campaign with differences in effectiveness (Supplementary Table 1); however, deaths and hospitalizations continued to decline, indicating that this vaccine mosaic gave reasonable protection in the Mexican population and decreased the severity of the registered cases. Nonetheless, another factor that helped to reduce the number of deaths as the pandemic continued was the acquisition of knowledge in treating the disease by health professionals. At the beginning of the pandemic, authorities recommended staying home until symptoms such as fever or chills and shortness of breath appeared. Currently, the recommendation is to receive health care if someone is suspected to be infected with SARS-CoV-2. This last recommendation leads to a better diagnosis and early treatment. Additionally, the introduction of antivirals and steroids, known for preventing progression to respiratory failure and death (20), were important factors in decreasing the death rate. It is important to highlight that a better treatment regimen in light of the molecular evolution of the virus has altered how the immune system faces the disease (21). In the fourth wave, the higher exposure of individuals to the new and more transmissible Omicron variant (due to the resumption of social activities) may have caused the observed upturn of CCs in the 18–49 age group. Additionally, in the case of the 50–59 age group, the loss of vaccine effectiveness caused by a decline in neutralizing antibodies against SARS-CoV-2 (22) may have been another reason for the increase in CCs, since these groups received the first two doses between May and July of 2021. Interestingly, W4 presented a decline in the frequency of CCs with comorbidities compared with the first wave (Supplementary Figure 4). This result can be associated with advances in vaccination and changes in the severity of the illness.

Fragile population

Worldwide reports have shown that while both sexes show the same susceptibility to COVID-19 infection, males belong to the population most vulnerable to COVID-19. Furthermore, aging and underlying comorbidities represent two serious risk factors for developing severe disease (23, 24). Consistent with those findings, in Mexico, while there was no difference in the likelihood of becoming infected between sexes (Table 1), males had higher odds than females of being hospitalized and dying (Tables 3, 4), and patients older than 50 years showed the highest odds of being hospitalized or dying compared to younger people (0–29 years). The presence of comorbidities also represented an important risk factor for the development of severe infection, increasing the odds of hospitalization and death by almost 2.5 times (Tables 3, 4). The highest risks of hospitalizations and deaths observed, especially in the first and second waves, could be related to the high prevalence of obesity, diabetes, and hypertension in all age groups of the Mexican population (25). In 2016, Mexico declared an obesity health emergency, where 76.0% of adults were overweight and obese[10] In 2020, the Health and Nutrition National Survey (ENSANUT as its acronym in Spanish) reported a diabetes prevalence of 15.7% and a prevalence of hypertension of 30.2% among people over 20 years of age[11] The combination of hypertension and diabetes strongly compromises the prognosis of COVID-19 patients. These conditions combined with aging represented the higher risk of death among those included in our study.

Limitations and strengths of the study

Especially in the first wave, the data suffer from bias due to undetected patients. However, the introduction of rapid testing allowed more complete monitoring of the pandemic. Nevertheless, the data are a highly reliable and complete source of information on the health strategy followed by the Mexican government. The experience of the first 2 years of the pandemic could help to define health policies for the follow-up of future epidemics and pandemics. It would be advisable to include the establishment of active contact system tracing in the national pandemic plan and defining a minimum threshold for the number of intensive care units at the regional level based on the population age, health and density. The information delivered by these data and their analysis could provide the general population with educational tools and access to health care services that improve their quality of life and allow them

10 https://www.gob.mx/salud/prensa/emite-la-secretaria-de-salud-emergencia-epidemiologica-por-diabetes-mellitus-y-obesidad

11 https://ensanut.insp.mx

to face this and subsequent epidemics as a healthy and informed population.

Data availability statement

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

Author contributions

AL, RW-C, M-EJ-C, and RG-R conceived and designed the methodology and analysis. RC, DP, and SZ contributed to data processing. AL and RG-R wrote the manuscript. AL, SZ, and RG-R prepared the figures and tables. RW-C, M-EJ-C, SZ, SL, RC, DP, RG-L, PI, BT, MR, CB, AH-E, NM, XR-G, JM-M, AS-L, AS-F, and JV-P reviewed the final version of the manuscript. CA helped improve the project, coordinated genomic surveillance, and reviewed the final version of the manuscript. RG-R coordinates the epidemiologic analysis group. All authors contributed to the article and approved the submitted version.

Funding

This research was partially funded by PAPIIT-DGAPA IN202821, awarded to RG-R and by grant "ANRS-Emerging Infectious Diseases. Project ECTZ184596: Combining mathematical modeling and phylodynamics analyses to characterize interactions between SARS-CoV-2 lineages within the increasingly vaccinated population of Mexico City awarded to CA. This is a contribution of the CoViGen-Mex [Consorcio Mexicano de Vigilancia Genómica, 2022].

Acknowledgments

We thank Juan Manuel Hurtado and Shirley Ainsworth for technical assistance. We also thank Dr. Dwight D. Leal for providing us with the SISVER data. We thank our editor for the suggestions made during the reviewing process that helped to improve the manuscript. AL was supported by a postdoctoral fellowship from the CONACyT project "Vigilancia Genómica del Virus SARS-CoV-2 en México" (PP-F003). RG-L is a recipient of a 2021 postdoctoral fellowship from CONACYT linked to ProNacEs #I1000/023/2021: C-08/2021.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.1050673/full#supplementary-material

References

1. Wang C, Horby PW, Hayden FG, Gao GF, A. novel coronavirus outbreak of global health concern. *Lancet*. (2020) 395:470–3. doi: 10.1016/S0140-6736(20)30185-9

2. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: A model-based analysis. *Lancet Infect Dis*. (2020) 20:669–77.

3. Gao YD, Ding M, Dong X, Zhang JJ, Kursat Azkur A, Azkur D, et al. Risk factors for severe and critically ill COVID-19 patients: A review. *Allergy Eur J Allergy Clin Immunol*. (2021) 76:428–455. doi: 10.1111/all.14657

4. Kuldeep D, Khan S, Tiwari R, Sircar S, Bhat S, Malik YS, et al. Coronavirus Disease 2019 –COVID-19. *Clin Microbiol Rev*. (2020) 33:1–48. doi: 10.1128/CMR.00028-20

5. Xu XW, Wu XX, Jiang XG, Xu KJ, Ying LJ, Ma CL, et al. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-CoV-2) outside of Wuhan, China: Retrospective case series. *BMJ*. (2020) 368:1–7. doi: 10.1136/bmj.m606

6. Mahase E. Covid-19: Runny nose, headache, and fatigue are commonest symptoms of omicron, early data show. *BMJ*. (2021) 373:n1654. doi: 10.1136/bmj.n1654

7. Boukadida C, Taboada B, Escalera-Zamudio M, Isa P, Ramírez-González JE, Vazquez-Perez JA, et al. Genomic characterization of SARS-CoV-2 isolated from patients with distinct disease outcomes in Mexico. *Microbiol Spectr*. (2022) 10:1–12. doi: 10.1128/spectrum.01249-21

8. Rodríguez-Maldonado AP, Vázquez-Pérez JA, Cedro-Tanda A, Taboada B, Boukadida C, Wong-Arámbula C, et al. Emergence and spread of the potential variant of interest (VOI) B.1.1.519 of SARS-CoV-2 predominantly present in Mexico. *Arch Virol*. (2021) 166:3173–77. doi: 10.1007/s00705-021-05208-6

9. Taboada B, Zárate S, García-López R, Muñoz-Medina JE, Sanchez-Flores A, Herrera-Estrella A, et al. Dominance of three sublineages of the SARS-CoV-2 Delta variant in Mexico. *Viruses*. (2022) 14:1165. doi: 10.3390/v14061165

10. Brauer F, Castillo-Chavez C. *Mathematical Models in Population Biology and Epidemiology*. New York: Springer-Verlag. (2001). doi: 10.1007/978-1-4757-3516-1

11. Ferrante P. The first 2 years of COVID-19 in Italy: Incidence, lethality, and health policies. *Front Public Heal.* (2022) 10. doi: 10.3389/fpubh.2022.986743

12. Suárez V, Quezada MS, Ruiz SO, de Jesús ER. Epidemiología de COVID-19 en México: del 27 de febrero al 30 de abril de 2020 [Epidemiology of COVID-19 in Mexico: from the 27th of February to the 30th of April 2020]. *Rev Clin Esp.* (2020) 220:463–71. doi: 10.1016/j.rce.2020.05.007

13. Zárate S, Taboada B, Muñoz-Medina JE, Iša P, Sanchez-Flores A, Boukadida C, et al. The alpha variant (B.1.1.7) of SARS-CoV-2 failed to become dominant in Mexico. *Microbiol Spectr.* (2022) 10:e0224021. doi: 10.1128/spectrum.02240-21

14. Khan NA, Al-Thani H, El-Menyar A. The emergence of new SARS-CoV-2 variant (Omicron) and increasing calls for COVID-19 vaccine boosters-The debate continues. *Travel Med Infect Dis.* (2022) 45:102246. doi: 10.1016/j.tmaid.2021.102246

15. Hu K, Lin L, Liang Y, Shao X, Hu Z, Luo H, et al. COVID-19: risk factors for severe cases of the Delta variant. *Aging (Albany NY).* (2021) 13:23459–70. doi: 10.18632/aging.203655

16. Meo SA, Meo AS, Al-Jassir FF, Klonoff DC. Omicron SARS-CoV-2 new variant: Global prevalence and biological and clinical characteristics. *Eur Rev Med Pharmacol Sci.* (2021) 25:8012–8. doi: 10.26355/eurrev_202112_27652

17. Wise J. Covid-19: Symptomatic infection with omicron variant is milder and shorter than with delta, study reports. *Bmj.* (2022) 2022:922. doi: 10.1136/bmj.o922

18. Palacio-Mejía LS, Hernández-Ávila JE, Hernández-Ávila M, Dyer-Leal D, Barranco A, Quezada-Sánchez AD, et al. Leading causes of excess mortality in Mexico during the COVID-19 pandemic 2020-2021: A death certificates study in a middle-income country. *Lancet Reg Health Am.* (2022) 13:100303. doi: 10.1016/j.lana.2022.100303

19. Lin L, Liu Y, Tang X, He D. The disease severity and clinical outcomes of SARS-CoV-2 variants of concern. *Front Public Heal.* (2021) 9:1–12. doi: 10.3389/fpubh.2021.775224

20. RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med.* (2021) 384:693–704. doi: 10.1056/NEJMoa2021436

21. Shah VK, Firmal P, Alam A, Ganguly D, Chattopadhyay S. Overview of immune response during SARS-CoV-2 infection: Lessons from the past. *Front Immunol.* (2020) 11:1–17. doi: 10.3389/fimmu.2020.01949

22. Liu J, Mao Q, Wu X, He Q, Bian L, Bai Y, et al. Considerations for the feasibility of neutralizing antibodies as a surrogate endpoint for COVID-19 Vaccines. *Front Immunol.* (2022) 13:1–14. doi: 10.3389/fimmu.2022.814365

23. Mukherjee S, Pahan K. Is COVID-19 Gender-sensitive? *J Neuroim Pharmacol.* (2021) 16:38–47. doi: 10.1007/s11481-020-09974-z

24. Bauer P, Brugger J, König F, Posch M. An international comparison of age and sex dependency of COVID-19 deaths in 2020: a descriptive analysis. *Sci Rep.* (2021) 11:1–11. doi: 10.1038/s41598-021-97711-8

25. Bello-Chavolla OY, Rojas-Martinez R, Aguilar-Salinas CA, Hernández-Avila M. Epidemiology of diabetes mellitus in Mexico. *Nutr Rev.* (2017) 75:4–12. doi: 10.1093/nutrit/nuw030

CITATION

Loza A, Wong-Chew RM, Jiménez-Corona M-E, Zárate S, López S, Ciria R, Palomares D, García-López R, Iša P, Taboada B, Rosales M, Boukadida C, Herrera-Estrella A, Mojica NS, Rivera-Gutierrez X, Muñoz-Medina JE, Salas-Lais AG, Sanchez-Flores A, Vazquez-Perez JA, Arias CF and Gutiérrez-Ríos RM (2023) Two-year follow-up of the COVID-19 pandemic in Mexico. *Front. Public Health* 10:1050673. doi: 10.3389/fpubh.2022.1050673

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 06 March 2023
DOI 10.3389/fpubh.2023.1087580

OPEN ACCESS

EDITED BY
Theodore Gyle Lewis,
Naval Postgraduate School, United States

REVIEWED BY
Elisabeth L. Zeilinger,
Medical University of Vienna, Austria
Rezvan Hosseinzadeh,
Babol University of Medical Sciences, Iran
Maria Francesca Piazza,
University of Genoa, Italy

*CORRESPONDENCE
Yvette Montcho
✉ yvettomontcho@gmail.com

†These authors have contributed equally to this work and share last authorship

SPECIALTY SECTION
This article was submitted to Infectious Diseases: Epidemiology and Prevention, a section of the journal Frontiers in Public Health

RECEIVED 02 November 2022
ACCEPTED 14 February 2023
PUBLISHED 06 March 2023

CITATION
Montcho Y, Klingler P, Lokonon BE, Tovissodé CF, Glèlè Kakaï R and Wolkewitz M (2023) Intensity and lag-time of non-pharmaceutical interventions on COVID-19 dynamics in German hospitals. *Front. Public Health* 11:1087580. doi: 10.3389/fpubh.2023.1087580

Intensity and lag-time of non-pharmaceutical interventions on COVID-19 dynamics in German hospitals

Yvette Montcho[1,2]*, Paul Klingler[2], Bruno Enagnon Lokonon[1,2], Chénangnon Frédéric Tovissodé[1], Romain Glèlè Kakaï[1†] and Martin Wolkewitz[2†]

[1]Laboratoire de Biomathématiques et d'Estimations Forestières, Université d'Abomey-Calavi, Cotonou, Benin, [2]Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany

Introduction: Evaluating the potential effects of non-pharmaceutical interventions on COVID-19 dynamics is challenging and controversially discussed in the literature. The reasons are manifold, and some of them are as follows. First, interventions are strongly correlated, making a specific contribution difficult to disentangle; second, time trends (including SARS-CoV-2 variants, vaccination coverage and seasonality) influence the potential effects; third, interventions influence the different populations and dynamics with a time delay.

Methods: In this article, we apply a distributed lag linear model on COVID-19 data from Germany from January 2020 to June 2022 to study intensity and lag time effects on the number of hospital patients and the number of prevalent intensive care patients diagnosed with polymerase chain reaction tests. We further discuss how the findings depend on the complexity of accounting for the seasonal trends.

Results and discussion: Our findings show that the first reducing effect of non-pharmaceutical interventions on the number of prevalent intensive care patients before vaccination can be expected not before a time lag of 5 days; the main effect is after a time lag of 10–15 days. In general, we denote that the number of hospital and prevalent intensive care patients decrease with an increase in the overall non-pharmaceutical interventions intensity with a time lag of 9 and 10 days. Finally, we emphasize a clear interpretation of the findings noting that a causal conclusion is challenging due to the lack of a suitable experimental study design.

KEYWORDS

lag-time effects, non-pharmaceutical interventions, distributed lag linear model, COVID-19 dynamics, Germany

Introduction

The coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (1), entered the world unexpectedly in 2019, dramatically changing human life (2). Infection occurs through respiratory droplets and contact routes during the incubation period. Outbreaks of the disease first appeared in Wuhan, Hubei Province, China (3), then spread to the United States, Europe, and Asia, reaching all continents. Since the World Health Organization (WHO) declared the disease a pandemic on March 11, 2020 (4), as of June 27, 2022, there have been more than 547 million confirmed cases worldwide and more than 6 million reported deaths (5). However, many

confirmed cases required hospitalization for several weeks while others require Intensive Care Unit (ICU) treatment (6). Due to a limited number of hospital beds, mainly ICU beds, many countries have adopted early control measures to prevent viral transmission and to avoid overloading the healthcare system (7, 8). Germany, the largest economic producer in Europe, has also inevitably experienced this pandemic. The first confirmed COVID-19 case in Germany was reported in late January 2020 following contact with an infected colleague from China (9). Afterwards, as of April 17, 2020, the Robert Koch-Institute (RKI) counts over 130,000 confirmed infections and about 4,000 deaths in Germany (10). To anticipate the massive flow of COVID-19, the federal government introduces public closures by closing public spaces such as schools, universities, and restaurants. Additional measures such as the national curfew ban and restrictions on people gatherings were also applied. In principle, people were advised to stay home as long as possible and leave home only for basic needs (11).

Several mathematical and statistical approaches have been developed to investigate the effectiveness of NPIs. Among these studies, Brauner et al. (12) applied a semi-mechanistic hierarchical Bayesian model to estimate the impact of NPIs on the time reproduction numbers in 41 countries during the first wave of the pandemic. They found that closing all educational institutions, limiting gatherings to 10 people or less, and closing face-to-face businesses reduced transmission considerably. The additional effect of stay-at-home orders was comparatively small. Nader et al. (13) used a non-parametric machine learning model to assess the effects of NPIs in relation to how long they have been in place and the effectiveness of NPIs in relation to their implementation date on the daily growth rate (relative increase in cumulative confirmed COVID-19 cases from 1 day to the next). They found that the closure and regulation of schools was the most important NPI, associated with a pronounced effect about 10 days after implementation. Sharma et al. (14) considered a semi-mechanistic hierarchical Bayesian model to examine the effect estimates for individual NPI during Europe's second wave of COVID-19 on daily cases and deaths. They concluded that business closures, educational institution closures, and gathering bans reduced transmission but less than they did during the first wave. Ge et al. (15) used a Bayesian inference model to assess the changing effect of NPIs and vaccination on reducing COVID-19 transmission based on the time reproduction numbers. Their results demonstrate that NPIs were complementary to vaccination in an effort to reduce COVID-19 transmission, and the relaxation of NPIs might depend on vaccination rates, control targets, and vaccine effectiveness concerning extant and emerging variants.

All the studies cited above have shown the effectiveness of NPIs. However, they did not consider the delay effects related to the NPIs implementation. A health effect is frequently associated with protracted exposures of varying intensity sustained in the past (16). The effects of exposing a particular event may not always be limited to the time of its occurrence and may appear with lag times (17). Policy lags are generally understood as unavoidable time delays. While there may exist several possible reasons for a lag, there is no general agreement on its length (18). This can be explained by the high sensitivity of the lagged and baseline exposure terms and also the implication of time-varying confounding variables in the models (19). Similar time lags have been noticed during the COVID-19 outbreak when various non-pharmaceutical interventions (NPIs) were implemented. Different policies may have different levels of effectiveness on disease spread, and the response to these policies is still unclear (20).

The main complexity of modeling and interpreting such phenomena lies in the additional temporal dimension needed to express the association, as the risk depends on both the intensity and timing of past exposures. This type of dependency is defined here as NPIs intensity-lag-response (Hospitalized cases and ICU cases) association. Statistical regression models are used to determine the relationship between predictors and outcomes and then estimate their effects. The Distributed Lag Model (DLM) models the exposure–response relationship and then introduces a series of consequences caused by this exposure to events. In addition, the method is used as well to determine the distribution of the subsequent effects after the occurrence of events (in lag times). This method has been developed for time series data and used in studies of various designs and data structures, cohort, case-control, or longitudinal studies (17). Distributed Lag Models have been successfully applied in epidemiological research (21–24).

Changes in the coronavirus infection dynamic in Germany led to the implementation of a policy like NPIs. The effect of NPIs may not be immediate since NPIs need some time to affect the pandemic situation. Then, it is reasonable to use the time lag concept in the analysis of COVID-19 research. This work aims to study intensity and lag time effects on the numbers of COVID-19 hospital and prevalent intensive care patients diagnosed with polymerase chain reaction tests in Germany from 10 January 2020 to June 2022. In this study, we applied a DLM to the number of COVID-19 hospital patients (Hospitalized cases) in Germany and the number of COVID-19 prevalent intensive care patients (ICU cases), considering all non-pharmaceutical interventions implemented in Germany and estimated their delayed effects. The results provide policymakers with essential information to make more informed decisions, considering the effect of NPIs, and their lag time in managing possible future pandemics.

Methods

Data description

Real data about the number of hospital patients (Hospitalized cases), prevalent intensive care patients (ICU cases), overall non-pharmaceutical intervention intensity (*NPIs*) for each of the 16 German states, the proportion of people who received at least two doses ($V2$) in Germany from January 2020 to June 2022, were extracted from the corona Daten platform site (https://www.corona-datenplattform.de). The first two variables (i.e., Hospitalized and ICU cases) were summed over states to obtain German countrywide data for our analysis.

The overall intensity of non-pharmaceutical interventions for Germany was computed using

$$NPIs_t = \sum_{f=1}^{16} \Im_f NPI_{f,t} \quad (1)$$

where $\Im_f$ is state f relative population share in Germany, and $NPI_{f,t}$ is the intensity of all NPIs implemented in the state f at time t. The whole German data was split into two sets to obtain a dataset before vaccination (from January 2020 to December 2020 data), a dataset with vaccination (from December 2020 to June 2022 data), and the entire dataset (from January 2020 to June 2022).

Distributed lag generalized linear model based regression

To assess the lag time effects of non-pharmaceutical interventions on COVID-19 dynamics, we used a DLM on each of the three datasets, considering the daily number of hospital patients (Hospitalized cases) and the number of prevalent intensive care patients (ICU cases) as response variables and overall NPIs intensity as exposure (predictor). Only time (variable date) is considered as a confounding variable for the data before vaccination. In addition to the time, the proportion of people who received at least two doses ($V2$) was considered a confounding variable for the data with vaccination and the entire data. The analyses were conducted within the statistical environment R version 4.0.3 (25) using the package **dlnm** (26).

Mathematically, a general model for time series data of outcomes y_t at time t can be written as:

$$g(\mu_t) = \alpha + \sum_{j=1}^{J} s_j\left(x_{tj}; \boldsymbol{\beta}_j\right) + \sum_{k=1}^{K} \phi_k u_{tk}, \tag{2}$$

where $\mu_t = \mathrm{E}\left(y_t\right)$ is the expected response for the day t, g is a monotonic link function (here $g = log$), and y_t $(t = 1, \ldots n)$ arises from a time series which is assumed to have an exponential family distribution (27). The function s_j is a smoother of the relationships between the variables x_j and the linear predictor, expressed by the parameter $\boldsymbol{\beta_j}$. Lastly, the u_k variables include other predictors with effects specified by the related coefficients ϕ_k. In this paper, we considered a set of variables x, which is overall NPIs intensity ($NPIs$) and two sets of variables u_1 and u_2, which are, respectively, date (t) and the proportion of people who received at least two doses ($V2$). We did a transformation to use nonlinear influences of the variable date (t) and captured its effect changing over time. This transformation is described in matrix notation as

$$f(t, \alpha) = z_t \alpha$$

where z_t is the t^{th} row of the matrix $\boldsymbol{Z}$. The transformation on a variable date (time) consists of using $ns(t, df)$, where df corresponds to the degree of freedom and ns, the natural spline function. The parameters for the natural spline are implicitly captured in the ns function of the R package *splines*. The matrix $\boldsymbol{Z}$ is generated automatically, and the parameters for the natural spline are implicitly captured by the function ns.

For the models considered, we assumed the influence of $NPIs$ and the proportion of people who received at least two doses ($V2$) to be linear, with no basis transformation. We assumed this since, from a preliminary investigation based on the Akaike Information Criterion (AIC), the linear model outperforms the non-linear model. The general notation for exposure-response linear relationships could be

$$s(x_t; \boldsymbol{\beta}) = \sum_{\ell=0}^{L} \beta_\ell x_{t-\ell} \tag{3}$$

where $\ell \in [0, L]$ is the lag duration, L (here $L = 30$) is the lag period over which exposure to x is assumed to affect the count change at time t, $x_{t-\ell}$ is exposure intensity at time $t-\ell$, and β_ℓ is the contribution from a unit increase in exposure x occurring at time $t - \ell$ in the past to a given count change measured at time t. For a more detailed description of the general theory on time-lagged models, we refer the reader to Gasparrini (26). The models are specified as indicated in Table 1 for the three datasets, along with the maximum number of degrees of freedom (df max) considered.

To implement delayed effects, the variables $NPIs$, $V2$, and date (t) are used to predict the two response variables (Hospitalized cases, ICU cases). The analysis is based on the models in Table 1, fitted through a generalized linear model with the Quasi-Poisson family, with natural splines of time with different degrees of freedom ($df = 1$ to df max) to describe long-time trends. A comparison was made between models with varying numbers of degrees of freedom using modified Akaike information criterion for models with overdispersed responses fitted through quasi-likelihood (28), given by:

$$\mathrm{QAIC} = -2\mathscr{L}(\hat{\boldsymbol{\theta}}) + 2\hat{\phi}k, \tag{4}$$

where k is the number of parameters, whereas $\mathscr{L}$ is the log-likelihood of the fitted model with parameters $\hat{\boldsymbol{\theta}}$ and $\hat{\phi}$, the estimated overdispersion parameter. Minimizing this criterion has led to the best model.

Sensitivity analyses were conducted to assess the impact of choices regarding the number of degrees of freedom (df) of the models. Specifically, we examine changes in the estimated overall effect associated with varying df for specifying the date or seasonal trend.

Results

Results for simple DLMs, assuming a linear relationship between response variables (number of hospital patients and number of prevalent intensive care patients) and all non-pharmaceutical interventions implemented in Germany (data before vaccination), and the proportion of people who received at least two doses (data with vaccination and entire Germany data) with a maximum lag equal to 30 days are summarized as follows. The Quasi AIC values for the number of degrees of freedom, df = 1 to df max (Table 1), are shown in Supplementary Figure S6. When used to compare different modeling choices with varying numbers of degrees of freedom, the Quasi AIC led to a parsimonious model, with 7 df for the data before vaccination (Supplementary Figures S7A, D), 15 df for the data with vaccination (Supplementary Figures S7B, E), 23 df for the whole German COVID-19 Hospitalized cases (Supplementary Figure S7C), and 19 df for whole German COVID-19 ICU cases (Supplementary Figure S7F) to describe the overall effect of exposures-lag on response variables.

10.3389/fpubh.2023.1087580

TABLE 1 Summary of model features. *df* max is the maximum degree of freedom considered.

Dataset	Model	*df* max
Data before vaccination	$g(\mu_t) = \alpha + ns(t, df) + \sum_{\ell=0}^{L} \beta_\ell NPIs_{t-\ell}$	10
Data with vaccination	$g(\mu_t) = \alpha + ns(t, df) + \sum_{\ell=0}^{L} \beta_{1\ell} NPIs_{t-\ell} + \sum_{\ell=0}^{L} \beta_{2\ell} V2_{t-\ell}$	15
Entire data		25

An overall graph of the effect of NPIs on the number of hospital patients (Hospitalized cases) and the number of prevalent intensive care patients (ICU cases) is provided in Figure 1, showing heat maps of the relative count change (RCC) of response variable along overall NPIs intensity and lags. Overall, the figure indicates that NPIs have an effect on the response variables before the vaccination program (Figures 1A, D) than on the response variables in the other two datasets (data with vaccination and the entire Germany data). The effect of NPIs is somewhat more immediate on Hospitalized cases before vaccination than on ICU cases. From 40% to 60% overall NPIs intensity, the mean relative count change of hospitalization before vaccination decreased from 1 to 0.85. In addition, a delay of 5 days was observed in the effect of overall NPIs on the ICU cases before vaccination, with a relative count change of 0.85 from 45% overall NPIs intensity.

Concerning data with vaccination and the entire German data, the lag time effects of non-pharmaceutical interventions on the number of hospital patients (Hospitalized cases) are immediate. However, the relative count change in hospitalization (data with vaccination) is high between lags 9 and 10 days from 55% to 60% overall NPIs intensity in Germany (see Figure 1B). The maximum effect of all non-pharmaceutical interventions implemented in Germany on Hospitalized and ICU cases during the vaccination programme is reached approximately at lag 25–30 days at 45–60% overall NPIs intensity.

Figure 2 presents the results from the sensitivity analyses, showing the overall effect of all NPIs implemented in Germany, summing up the contributions for the 30 days of lag considered in the analysis. There was an overall decrease in the number of patients in hospital and intensive care units with increasing overall NPIs intensity. This relative count change (RCC) was canceled out for the data before vaccination and reached its minimum value of around 0.3 for the data with vaccination and the entire data at 55–60% overall NPIs intensity.

Discussions

This study used a Distributed Lag Linear Model (DLM) to evaluate the lag time effects of non-pharmaceutical interventions on the number of COVID-19 hospital patients and the number of prevalent COVID-19 intensive care patients. Based on the results of this analysis, it is important to investigate both the lag and magnitude of NPIs impact jointly (17).

An epidemiological discussion of DLM choice emphasized the need to balance capturing detail and allow for interpretation (23). Despite its conceptual simplicity, DLM specifications allow for a wide range of models, including simple previously used models and more complex variants. There is a challenge in choosing between alternatives when there is an abundance of choices (number of degrees of freedom, maximum lag). We used quasi-Akaike information criteria to guide the selection of the number of degrees of freedom (*df*) of the variable *time*. Due to the lack of consensus about what constitutes an optimal model, sensitivity analyses are particularly important for assessing how key conclusions depend on the model's number of degrees of freedom.

The results of our study revealed that NPIs have a positive effect on the number of hospital patients (Hospitalized cases) and of prevalent intensive care patients (ICU cases) for all the datasets (data before vaccination, data with vaccination or the entire COVID-19 German data) since the overall non-pharmaceutical intervention decrease the number of incident hospital patients (Hospitalized cases) and the number of prevalent intensive care patients (ICU cases). These results are similar to a previous study which showed that some interventions are effective in reducing the advent of the pandemic (29). We found that the first reducing effect of NPIs on the number of prevalent intensive care patients before vaccination cannot be expected before a time lag of 5 days. As 5 days seems to be a short delay effect, it might be possible that already the announcement of planned NPI introductions from policy makers have an impact on the social behavior such as contact activities and hence on the pandemic dynamics. However, our results also suggest that the main effect is after a time lag of 10–15 days.

However, the results contrast with a previous study which evaluated NPIs effects on the COVID-19 pandemic in Germany and three other European countries using the Granger Causality test (30). In addition, previous studies have focussed on the effect of NPIs on infectious cases and death (11, 14), recovered cases (11), daily growth rate (13), or time-varying reproduction number (12, 15) contrary to this study. The contradiction could be due to the fact that the previous study took into account the effect of NPIs on the number of infections in the second wave, whereas we evaluated the way how NPIs are associated with a decrease in the number of patients in hospitals and intensive care units diagnosed with PCR in several waves. We note that the number of infections in the general population usually depends on several non-infection-related factors, such as testing behavior and the day of the week and is therefore often subject to reporting bias and delays. Thus, this outcome is less specific and prone to higher statistical noise than the number of patients in hospital and intensive care units. In addition to the DLM analyses, the Granger causality test methodology has been applied to our data set. They resulted in a similar conclusion, even though all-time series have to be transformed by the second differences to reach stationarity and decomposed with respect to time trends and seasonality (data not shown).

The advantage of applying DLM is that it is flexible and provides a comprehensive scheme for interpreting outcomes from exposure-lag-response associations contrary to other statistical

10.3389/fpubh.2023.1087580

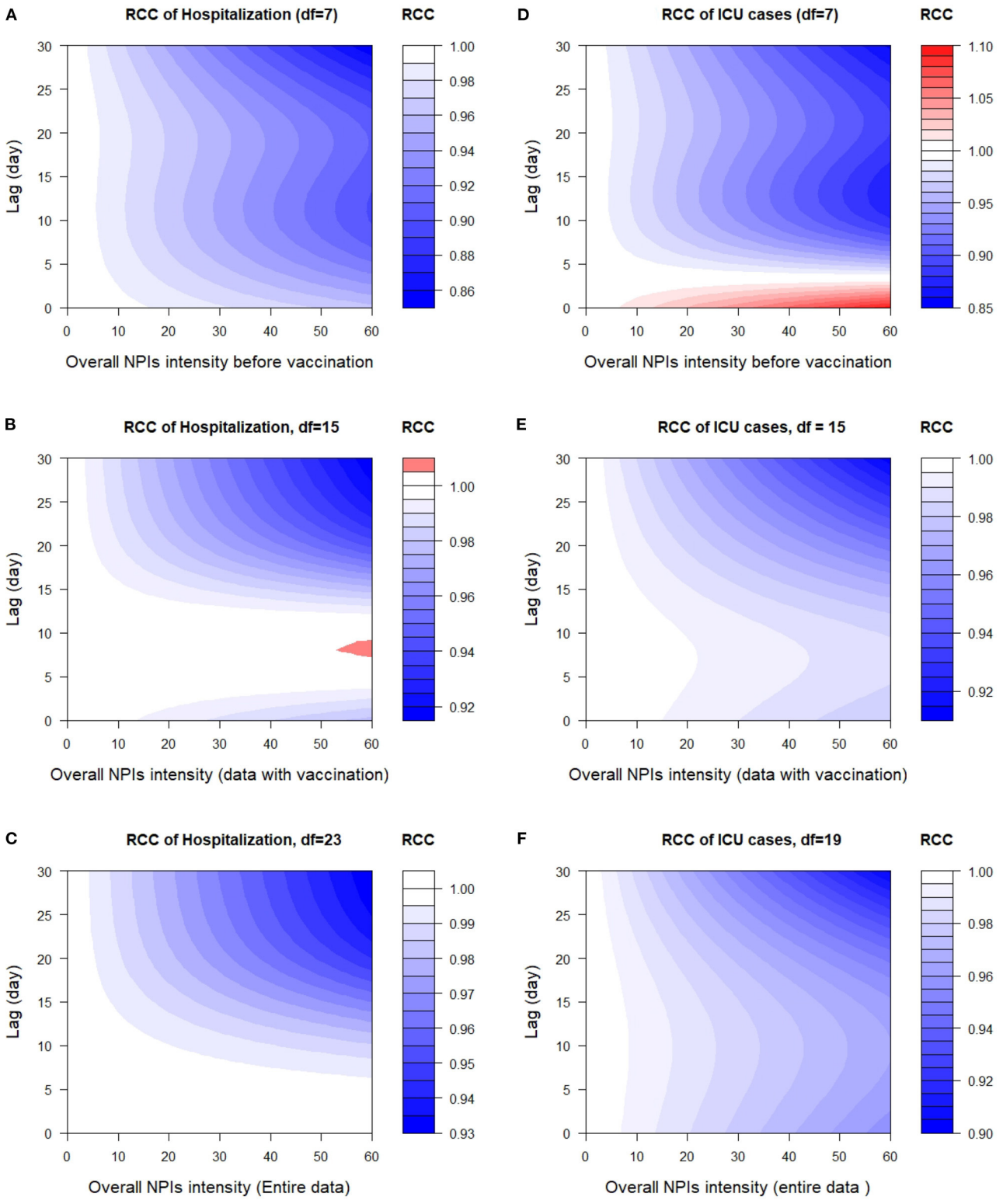


FIGURE 1
2-D plot of relative count change (RCC) along NPIs and lags on the number of hospital patients (Hospitalized cases) and the number of prevalent intensive care patients (ICU cases). **(A)** RCC of Hospitalization (df = 7), **(B)** RCC of Hospitalization, df = 15, **(C)** RCC of Hospitalization, df = 23, **(D)** RCC of ICU cases (df = 7), **(E)** RCC of ICU cases, df = 15, and **(F)** RCC of ICU cases, df = 19.

approaches (16). The main limitation of our analysis is that our results cannot—strictly speaking—be interpreted as causal effects; they are associations. To increase the ability to infer causality, pragmatic study designs such as the SMART (Sequential, Multiple-Assignment Randomized Trial), stepped wedge, and preference designs could have been an option (31). An interesting

10.3389/fpubh.2023.1087580

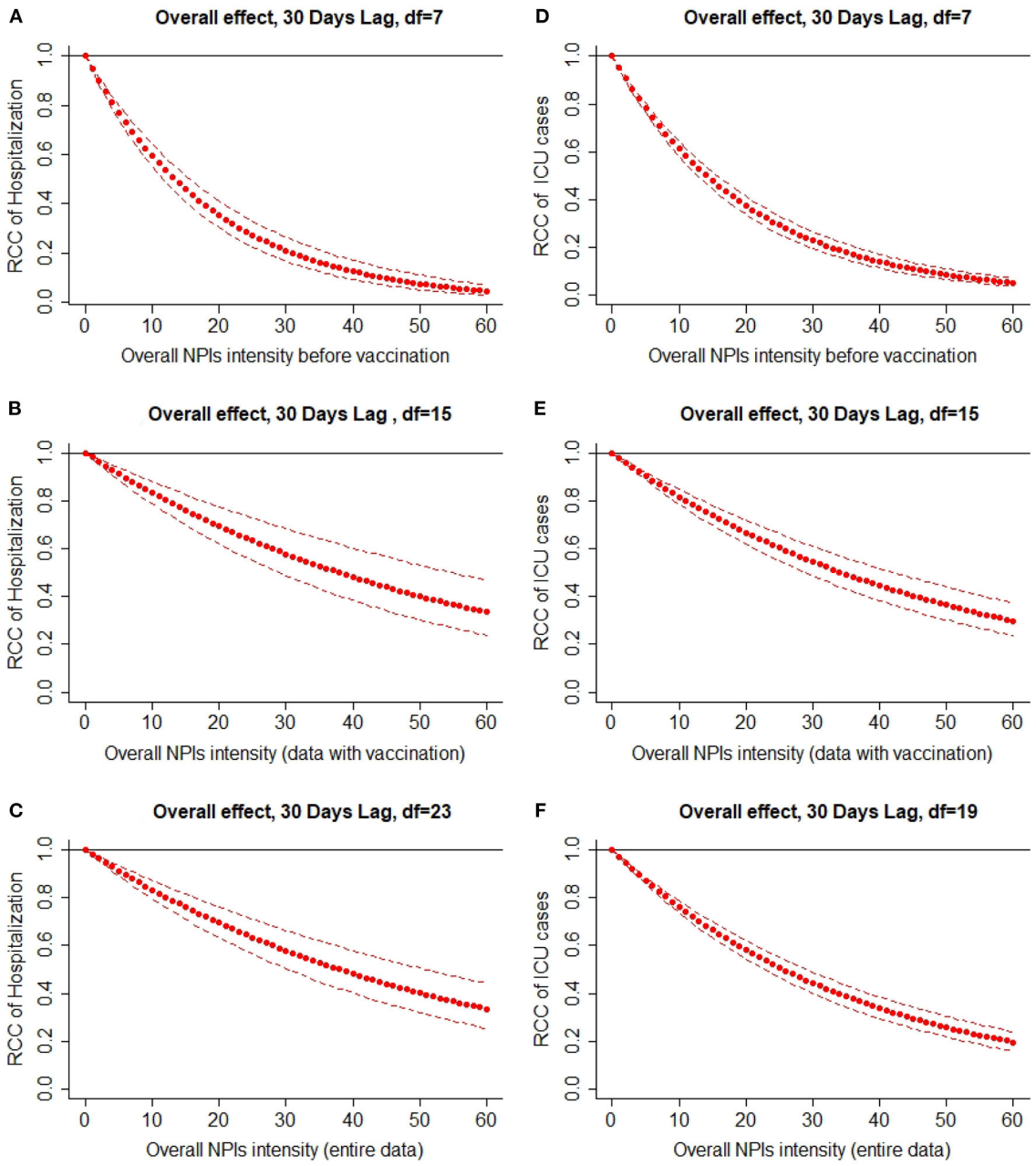


FIGURE 2
Overall lag effect of NPIs intensity on the number of incident hospital patients (Hospitalized cases) and the number of prevalent intensive care patients (ICU cases) with 95% confidence intervals. **(A)** Overall effect, 30 days lag, df = 7, **(B)** overall effect, 30 days lag, df = 15, **(C)** overall effect, 30 days lag, df = 23, **(D)** overall effect, 30 days lag, df = 7, **(E)** overall effect, 30 days lag, df = 15 and **(F)** overall effect, 30 days lag, df = 19.

design and analysis is also the trial emulation approach (32), where cluster-randomized trials are mimicked. However, in Germany, the introduction, timing, and intensity of NPIs were quite homogeneously distributed across Germany (see Supplementary material); hence, the above-mentioned designs were hardly feasible in practice.

Obtaining "zero Hospitalized or ICU COVID-19 cases" may not be achievable, but reducing the number to a level that can be managed by the health system might be a feasible goal. This paper considered a bundle or the overall intensity of NPIs implemented in Germany; an isolation of specific NPIs is hardly feasible due to high correlations. However, a society's ability to fight a pandemic can be influenced by various factors, including how well the public health care system works, how the government manages risk, transparency of information flow, how it is driven by politics, corporate and citizen compliance, etc. Consequently, further studies are needed to investigate what determines whether or not the COVID-19 pandemic is controlled.

An application to the COVID-19 data from Germany indicates that the Distributed Lag Model can be used to assess the

effect of control measures dictated by health policies with changes in the transmission dynamics of the studied disease. In addition, using them can assist policymakers in planning appropriate and timely strategies and allocating resources (20). Our results can inform political decision makers regarding how long NPIs need to be implemented to take effect on controlling the COVID-19 dynamics in hospitals. However, we focused only on this aspect. Beyond that, NPIs create tremendous economic and social collateral damages of multifaceted dimensions such as psychological long-term effects on mental health of children due to long school closures and contact distancing. Thus, political decision makers need to trade off NPIs effects on hospitals against the collateral damages of NPIs in the society.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.corona-datenplattform.de.

Author contributions

YM, RGK, and MW contributed to the study conception and design. YM carried out data analysis and wrote the initial manuscript draft. MW contributed to data collection, supervised data analysis, and revised the manuscript. RGK supervised data analysis and revised the manuscript. All authors read and revised the initial manuscript. All authors contributed to the article and approved the submitted version.

Funding

The authors acknowledge the support from the Humboldt Research Hub SEMCA, funded by the German Federal Foreign Office with the support of the Alexander von Humboldt Foundation (AvH).

Acknowledgments

We thank Nia Hermanutz for performing the additional analyses using the Granger causality test methodology.

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.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1087580/full#supplementary-material

References

1. Niu Y, Rui J, Wang Q, Zhang W, Chen Z, Xie F, et al. Containing the transmission of COVID-19: a modeling study in 160 countries. *Front Med.* (2021) 8:701836. doi: 10.3389/fmed.2021.701836

2. Khalifa SA, Swilam MM, El-Wahed AAA, Du M, El-Seedi HH, Kai G, et al. Beyond the pandemic: COVID-19 pandemic changed the face of life. *Int J Environ Res Public Health.* (2021) 18:5645. doi: 10.3390/ijerph 18115645

3. Spiteri G, Fielding J, Diercke M, Campese C, Enouf V, Gaymard A, et al. First cases of coronavirus disease 2019 (COVID-19) in the WHO European Region, 24 January to 21 February 2020. *Eurosurveillance.* (2020). 25:2000178. doi: 10.2807/1560-7917.ES.2020.25.9.2000178

4. Bialek S, Boundy E, Bowen V, Chow N, Cohn A, Dowling N, et al. Severe outcomes among patients with coronavirus disease 2019 (COVID-19)-United States, February 12–March 16, 2020. *MMWR Morbid Mortal Wkly Rep.* (2020) 69:343. doi: 10.15585/mmwr.mm6912e2

5. WHO. *World Health Organization Coronavirus (COVID-19)* (2022). Available online at: https://covid19.who.int/ (accessed August 13, 2022).

6. Phua J, Weng L, Ling L, Egi M, Lim CM, Divatia JV, et al. Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations. *Lancet Respir Med.*(2020) 8:506–17. doi: 10.1016/S2213-2600(20)30161-2

7. Ma X, Vervoort D. Critical care capacity during the COVID-19 pandemic: global availability of intensive care beds. *J Crit Care.* (2020) 58:96–7. doi: 10.1016/j.jcrc.2020.04.012

8. Cowling BJ, Aiello AE. Public health measures to slow community spread of coronavirus disease (2019). *J Infect Dis.* (2020) 221:1749–51. doi: 10.1093/infdis/jiaa123

9. Böhmer MM, Buchholz U, Corman VM, Hoch M, Katz K, Marosevic DV, et al. Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. *Lancet Infect Dis.* (2020) 20:920–8. doi: 10.1016/S1473-3099(20)30314-5

10. Barbarossa MV, Fuhrmann J, Meinke JH, Krieg S, Varma HV, Castelletti N, et al. Modeling the spread of COVID-19 in Germany: early assessment and possible scenarios. *PLoS ONE.* (2020) 15:e0238559. doi: 10.1371/journal.pone.0238559

11. Khairulbahri M. Modeling the flow of the COVID-19 in Germany: the cohort SEIR model based on the system dynamics approach. *medRXiv.* (2020). doi: 10.1101/2020.12.21.20248605

12. Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Gavenčiak T, et al. Inferring the effectiveness of government interventions against COVID-19. *Science.* (2021) 371:eabd9338. doi: 10.1126/science.abd9338

13. Nader IW, Zeilinger EL, Jomar D, Zauchner C. Onset of effects of non-pharmaceutical interventions on COVID-19 infection rates in 176 countries. *BMC Public Health.* (2021) 21:1472. doi: 10.1186/s12889-021-11530-0

14. Sharma M, Mindermann S, Rogers-Smith C, Leech G, Snodin B, Ahuja J, et al. Understanding the effectiveness of government interventions against the resurgence of COVID-19 in Europe. *Nat Commun.* (2021) 12:5820. doi: 10.1038/s41467-021-26013-4

15. Ge Y, Zhang WB, Wu X, Ruktanonchai CW, Liu H, Wang J, et al. Untangling the changing impact of non-pharmaceutical interventions and vaccination on European COVID-19 trajectories. *Nat Commun.* (2022) 13:3106. doi: 10.1038/s41467-022-30897-1

16. Gasparrini A. Modeling exposure–lag–response associations with distributed lag non-linear models. *Stat Med.* (2014) 33:881–99. doi: 10.1002/sim.5963

17. Entezari A, Mayvaneh F. Applying the distributed lag non-linear model (DLNM) in epidemiology: temperature and mortality in Mashhad. *Iran J Public Health.* (2019) 48:2108. doi: 10.18502/ijph.v48i11.3539

18. Jovanovski T, Muric M. The phenomenon of lag in application of the measures of monetary policy. *Econ. Res.* (2011) 24:154–63. doi: 10.1080/1331677X.2011.11517463

19. Mansournia MA, Naimi AI, Greenland S. The implications of using lagged and baseline exposure terms in longitudinal causal and regression models. *Am J Epidemiol.* (2019) 188:753–9. doi: 10.1093/aje/kwy273

20. Bian Z, Zuo F, Gao J, Chen Y, Venkata SSCP, Bernardes SD, et al. Time lag effects of COVID-19 policies on transportation systems: a comparative study of New York City and Seattle. *Transport Res A Policy Pract.* (2021) 145:269–83. doi: 10.1016/j.tra.2021.01.019

21. Analitis A, Katsouyanni K, Biggeri A, Baccini M, Forsberg B, Bisanti L, et al. Effects of cold weather on mortality: results from 15 European cities within the PHEWE project. *Am J Epidemiol.* (2008) 168:1397–408. doi: 10.1093/aje/kwn266

22. Anderson BG, Bell ML. Weather-related mortality: how heat, cold, and heat waves affect mortality in the United States. *Epidemiology.* (2009) 20:205–13. doi: 10.1097/EDE.0b013e318190ee08

23. Armstrong B. Models for the relationship between ambient temperature and daily mortality. *Epidemiology.* (2006) 17:624–31. doi: 10.1097/01.ede.0000239732.50999.8f

24. Schwartz J. The distributed lag between air pollution and daily deaths. *Epidemiology.* (2000) 11:320–6. doi: 10.1097/00001648-200005000-00016

25. Team RDC. *R: A Language and Environment lor Statistical Computing.* Vienna: R Foundation lor Statistical Computing (2020). Available online at: http://www.R-project.org/ (accessed August 15, 2022).

26. Gasparrini A. Distributed lag linear and non-linear models in R: the package dlnm. *J Stat Softw.* (2011) 43:1–20. doi: 10.18637/jss.v043.i08

27. Dobson AJ, Barnett AG. *An Introduction to Generalized Linear Models.* Boca Raton, FL: Chapman and Hall/CRC (2018).

28. Peng RD, Dominici F, Louis TA. Model choice in time series studies of air pollution and mortality. *J Roy Stat Soc A Stat Soc.* (2006) 169:179–203. doi: 10.1111/j.1467-985X.2006.00410.x

29. Tsuchiya Y, Yoshimura M, Sato Y, Kuwata K, Toh S, Holbrook-Smith D, et al. Probing strigolactone receptors in Striga hermonthica with fluorescence. *Science.* (2015) 349:864–8. doi: 10.1126/science.aab383

30. Gianino MM, Nurchis MC, Politano G, Rousset S, Damiani G. Evaluation of the strategies to control COVID-19 pandemic in four European countries. *Front Public Health.* (2021) 9:700811. doi: 10.3389/fpubh.2021.700811

31. Digitale JC, Stojanovski K, McCulloch CE, Handley MA. Study designs to assess real-world interventions to prevent COVID-19. *Front Public Health.* (2021) 9:657976. doi: 10.3389/fpubh.2021.657976

32. Ben-Michael E, Feller A, Stuart EA. A trial emulation approach for policy evaluations with group-level longitudinal data. *Epidemiology.* (2021) 32:533–40. doi: 10.1097/EDE.0000000000001369

frontiers | Frontiers in Public Health

TYPE Perspective
PUBLISHED 16 March 2023
DOI 10.3389/fpubh.2023.1111661

OPEN ACCESS

EDITED BY
Pierpaolo Ferrante,
National Institute for Insurance Against Accidents at Work (INAIL), Italy

REVIEWED BY
Andrey Zheluk,
Charles Sturt University, Australia

*CORRESPONDENCE
Shi Chen
✉ schen56@uncc.edu

SPECIALTY SECTION
This article was submitted to Infectious Diseases: Epidemiology and Prevention, a section of the journal Frontiers in Public Health

RECEIVED 29 November 2022
ACCEPTED 21 February 2023
PUBLISHED 16 March 2023

CITATION
Chen S, Yin SJ, Guo Y, Ge Y, Janies D, Dulin M, Brown C, Robinson P and Zhang D (2023) Content and sentiment surveillance (CSI): A critical component for modeling modern epidemics. *Front. Public Health* 11:1111661. doi: 10.3389/fpubh.2023.1111661

Content and sentiment surveillance (CSI): A critical component for modeling modern epidemics

Shi Chen[1,2,3]*, Shuhua Jessica Yin[4], Yuqi Guo[2,5], Yaorong Ge[4], Daniel Janies[6], Michael Dulin[1,3], Cheryl Brown[2,7], Patrick Robinson[1,3] and Dongsong Zhang[2,8]

[1]Department of Public Health Sciences, College of Health and Human Services, University of North Carolina at Charlotte, Charlotte, NC, United States, [2]School of Data Science, University of North Carolina at Charlotte, Charlotte, NC, United States, [3]Academy for Population Health Innovation, University of North Carolina at Charlotte, Charlotte, NC, United States, [4]Department of Software and Information Systems, College of Computing and Informatics, University of North Carolina at Charlotte, Charlotte, NC, United States, [5]School of Social Work, College of Health and Human Services, University of North Carolina at Charlotte, Charlotte, NC, United States, [6]Department of Bioinformatics and Genomics, College of Computing and Informatics, University of North Carolina at Charlotte, Charlotte, NC, United States, [7]Department of Political Science and Public Administration, College of Liberal Arts and Sciences, University of North Carolina at Charlotte, Charlotte, NC, United States, [8]Belk College of Business, University of North Carolina at Charlotte, Charlotte, NC, United States

Comprehensive surveillance systems are the key to provide accurate data for effective modeling. Traditional symptom-based case surveillance has been joined with recent genomic, serologic, and environment surveillance to provide more integrated disease surveillance systems. A major gap in comprehensive disease surveillance is to accurately monitor potential population behavioral changes in real-time. Population-wide behaviors such as compliance with various interventions and vaccination acceptance significantly influence and drive the overall epidemic dynamics in the society. Original infoveillance utilizes online query data (e.g., Google and Wikipedia search of a specific content topic such as an epidemic) and later focuses on large volumes of online discourse data about the from social media platforms and further augments epidemic modeling. It mainly uses number of posts to approximate public awareness of the disease, and further compares with observed epidemic dynamics for better projection. The current COVID-19 pandemic shows that there is an urgency to further harness the rich, detailed content and sentiment information, which can provide more accurate and granular information on public awareness and perceptions toward multiple aspects of the disease, especially various interventions. In this perspective paper, we describe a novel conceptual analytical framework of content and sentiment infoveillance (CSI) and integration with epidemic modeling. This CSI framework includes data retrieval and pre-processing; information extraction *via* natural language processing to identify and quantify detailed time, location, content, and sentiment information; and integrating infoveillance with common epidemic modeling techniques of both mechanistic and data-driven methods. CSI complements and significantly enhances current epidemic models for more informed decision by integrating behavioral aspects from detailed, instantaneous infoveillance from massive social media data.

KEYWORDS

infoveillance, modeling, behavior, parameterization, mechanism, data-driven (DD)

1. Introduction

Mathematical models, such as the mechanistic susceptible exposed infectious recovered (SEIR) type modeling paradigm and alternative data-driven methods, have made investigations on epidemics across the globe (1). Epidemic modeling can systematically characterize epidemiological processes (e.g., transmission, immunization, hospitalization, recovery, etc.) and provide key metrics for epidemic projection, intervention, and resource optimization. In order to achieve these goals, a fundamental layer in epidemic modeling is to ensure comprehensive, accurate, and effective data collection through surveillance systems. The grand challenge of current epidemic modeling is to effectively identify, integrate, and analyze heterogeneous, cross-scale, and multimodal data from pathogen biology, human cognition and behavior, to social determinants of health (2).

Currently, many surveillance systems, such as the U.S. National Notifiable Diseases Surveillance System (NNDSS), have been developed from reported symptomatic cases. Additional surveillance systems, including genomic, serologic, and environmental surveillance systems in the CDC COVID-19 data dashboard, have been developed across national, state, and local levels, along with many other regions in the world (3, 4).

A key driver of epidemic dynamics is host cognition and behavior, such as adherence to interventions and vaccine acceptance. However, effective monitoring of behavior continuously on a large scale is challenging, as is quantifying its relevance to the observed health outcomes in an epidemic. Traditional participatory survey-based surveillance cannot provide comprehensive and continuous characterization of public perceptions toward the epidemic and various interventions, especially vaccination. Accurate characterization of public perceptions at different locations during different phases of an epidemic is critical to our efforts in designing and evaluating targeted interventions. To address this major issue, infoveillance, which observes, retrieves, and analyzes public online discourse especially on social media, has been developed since the 2000's (5–8). Infoveillance is implemented to monitor many diseases, including seasonal and pandemic influenza, Ebola, and COVID-19 epidemics (9–11). Traditional infoveillance approaches analyze online discourse dynamics of health issues by counting relevant posts and/or search queries. For instance, using COVID-19 specific terms, daily number or percentage of COVID-19-related posts and search queries can be counted. The discourse dynamics, expressed as the time series of the absolute number or relative percentage of the disease, is then compared with important health outcomes such as reported case, vaccination uptake, hospitalization, and death. Studies have shown that effective infoveillance can help predict early surges of an epidemic (9–11).

Nevertheless, we argue that traditional infoveillance—albeit offering advances in surveillance of various disease outbreaks, timing, and locations– lacks detailed extraction and characterization of dynamic public awareness, perceptions and sentiments toward interventions, which reflect behavioral changes and drive epidemic dynamics. Traditional infoveillance focuses on time series of posting counts or queries of the health issue, and ignores the large amount and rich information embedded in the actual contents of these discourses. With more recent advances in natural language processing (NLP), it is possible to further extract important information, such as contents and sentiments from social media posts (12–17, 20, 21). In this perspective paper, we introduce a conceptual framework of comprehensive content and sentiment infoveillance (CSI), including data mining and knowledge discovery of content and sentiment from social media discussions on epidemics (especially toward important interventions such as vaccination) with spatio-temporal variations, and integration with existing mechanistic and data-driven epidemic modeling techniques.

2. Content and sentiment infoveillance framework for epidemic modeling

2.1. Data retrieval, sampling, and pre-processing

Online discourse data are retrieved and sampled *via* application process interfaces (APIs). Many online platforms, such as Google, Wikipedia, Twitter, Instagram, Facebook, TikTok, have a public API. For instance, COVID-19 Twitter discussion will be acquired *via* the Twitter API. Specific keywords and key phrases related to COVID-19 will be predetermined to the API query, along with other specifications such as frequency and rate of sampling. Because of the sheer volume of COVID-19 discussions, usually a daily random 1% sampling will pull millions tweets per day, adequate for further CSI. Raw data (usually in JSON file format) from API query consist of two components: post body, including mainly the textual data of the post; and post metadata, including posting time, location, ID information (ID, display name, verification status, number of friends, number of followers, etc.), and post virality measures (e.g., numbers of shares, replies, and likes). Raw JSON data are transformed into a dataframe for further mining and analyses. Each row in the dataframe corresponds to a specific tweet post, with both post body and metadata across multiple columns (Figure 1).

2.2. Data mining and natural language processing

Once raw data are retrieved and pre-processed, the major task is to transform the unstructured textual data into numeric format for effective analyses. We propose a standardized four dimensions

Abbreviations: ABM, Agent-Based Model; API, Application Process Interface; BERT, Bidirectional Encoder Representations from Transformers; LDA, Latent Dirichlet Allocation; LUT, Look Up Table; NNDSS, National Notifiable Diseases Surveillance System; NLP, Natural Language Processing; RNN, Recurrent Neural Network; SEIR, Susceptible, Exposed, Infected, Recovered; SVM, Support Vector Machine; TF-IDF, Term Frequency-Inverse Document Frequency.

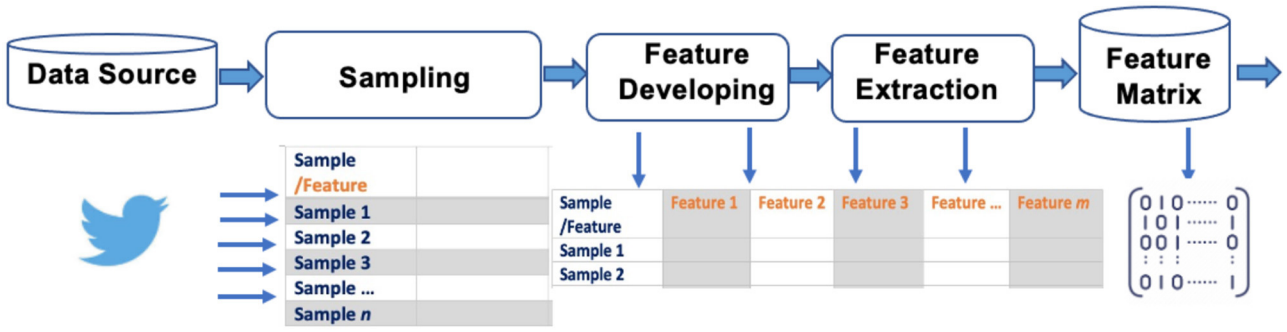


FIGURE 1
Social media data retrieval, sampling, and pre-processing. using Twitter as an example for online social platforms.

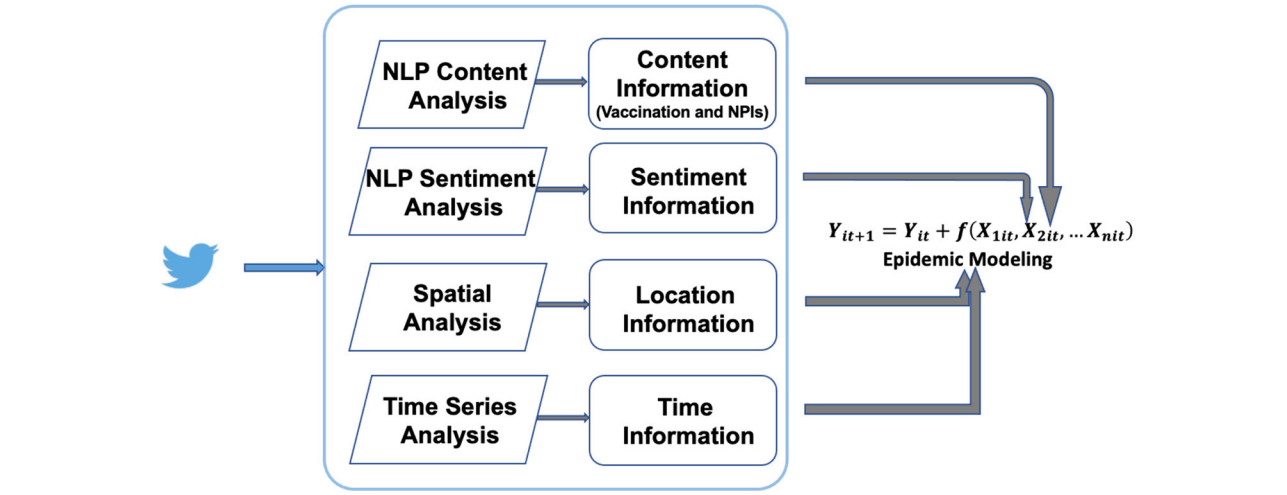


FIGURE 2
Social media data mining, NLP, and integration with epidemic modeling.

of information to be extracted from each post: (1) time, (2) location, (3) content, and (4) sentiment.

The first two dimensions, time and location will be derived from metadata. However, not all social media users allow sharing their locations, nor would location information always exist in a post. A post can be a general discussion of the health issue. Location data may be determined *via* natural language processing (NLP) of the post. A feasible solution is to develop a rule-based lookup table (LUT) with pre-defined location term database. Depending on the nature and scope of a specific study, the LUT may contain state-level (i.e., names, abbreviations, or other synonyms/indicators of the 51 states, DC, and oversea territories), or county-level (e.g., Mecklenburg County where Charlotte is located) terms. Then, a post is compared with the LUT to determine if there is a match of location terms. A large sample size during the initial API query will ensure adequate spatial coverage. Alternatively, if specific locations are of interest, these locations can be pre-specified in the initial API query (e.g., adding specific location keywords in the query) for sampling.

A third and perhaps the most important dimension is the content (also known as topic or narrative) of the post. A feasible approach is to use LUT with predefined terms to identify specific contents, similar to spatial information identification. For example, "vaccine," "vaccination," "inoculation, "shot," "jab," "immunization," "herd immunity" are all be relevant terms of vaccination contents. However, unlike spatial information which can be exhaustively captured by LUT, content information has much more variability and may include terms that are missed in the predetermined LUT. On the other hand, certain terms can have a low specificity. For instance, although "shot" is often interchangeable with "vaccination," a post mentioning "shot" may not be related to vaccination at all, causing a "false positive" sample of vaccination-related content.

Recent advances in NLP are able to accurately and comprehensively identify content information from textual data (e.g., a social media post, a sentence, a document, and a corpus with multiple documents). Latent Dirichlet allocation (LDA) is a probabilistic-based technique that generates clusters of distributions of words to identify latent topics from input

texts. The words in different topics are assumed to have Dirichlet distributions, hence the term LDA. Performance metrics of LDA include perplexity and topic coherence score, which evaluate model predictability and quality of topics, respectively. Outcomes of LDA are the most relevant words in each identified topic. Note that LDA is an unsupervised clustering algorithm, i.e., identified topics come unlabeled from LDA. Therefore, final interpretation and labeling of each topic requires domain knowledge from researchers.

Bidirectional encoder representations from transformers (BERT) is another emerging and powerful NLP technique for topic modeling. The textual data of posts are fed into BERT to generate different levels of embeddings based on the contexts of the word. BERT is constructed from deep neural networks (DNN) with millions of hyperparameters and pre-trained by massive corpus from online text sources including Wikipedia. BERT is able to learn high-level representations of textual data, and cluster reduced embeddings more effectively than probabilistic-based LDA. The clusters are then processed *via* term frequency-inverse document frequency (TF-IDF) to further create topics from clusters. Finally, similar to LDA, domain knowledge is applied to label and interpret identified contents.

In short, different NLP (LUT, LDA, BERT) all fulfill the same objective: further breaking down posts with textual data into more granular, specific contents for further analyses. Certain contents are specifically relevant for epidemic modeling, e.g., discussions on vaccinations and other interventions.

Lastly, sentiment analysis is carried out to evaluate sentiments and/or emotions in the post. Sentiments can be an important indicator of potential health behavioral change, which is crucial for epidemic processes such as infection and vaccination. Depending on the natural of the research, sentiment can be quantified as binary positive or negative, discrete scales (e.g., positive, neutral, or negative) or more granular Paul Ekman six emotion classification and more continuous emotion axes (20–23). Various methods can be used for sentiment analysis, including BERT and ML classification methods (e.g., support vector machine, SVM). In particular, sentiments toward interventions (NPIs and vaccination) can be critical indicators of changes in behavior and epidemic dynamics during the COVID-19 pandemic.

The post-specific dataframe (row as post and columns as the four major dimensions of information) will then be transformed into multiple specific dataframes based on posts' contents, for instance, vaccine-specific, mask mandate-specific, social distancing-specific. The conceptual analytical and NLP framework is presented in Figure 2.

2.3. Integrating novel CSI with epidemic modeling: a proposed case study for COVID-19

Once the four dimensions of information – time, location, content, and sentiment – have been retrieved from social media posts, we further recommend the following framework to integrate this novel CSI with epidemic modeling with a case study for COVID-19. This novel CSI significantly increases the amount of information from post contents and sentiments especially on public sentiments toward vaccination and other interventions during the pandemic. We will further extract intervention-specific content, along with sentiments toward these interventions and spatial information. For instance, we will construct a time series of vaccination-related posts (CV_{tj}) at a given location j. CV_{tj} can be either absolute number of posts, or relative percentage in all sampled posts at day t. In general, percentage of vaccination-related posts reflects public awareness of the content such as vaccination. The dynamic change of a specific content (e.g., vaccination) percentage reflects the varying degrees of public *awareness* during different phases of the pandemic. In addition, sentiment shifts of the vaccination content topic will also be captured by the sentiment time series, which can be expressed as the percentage of positive or negative sentiment toward vaccination, SV_{tj}. The sentiment time series reflects the dynamic change of vaccination *acceptance* by the public at the location j. For instance, vaccination acceptance can be evaluated by positive sentiments or emotions expressed in the posts. Similarly, positive sentiments toward other NPIs (e.g., social distancing, mask-wearing) may indicate increased willingness of compliance with these health policies. These detailed, dynamic characterizations of public awareness and acceptance of vaccination and other NPIs are critical indicators of health decisions and potential behavioral changes (e.g., actively seeking vaccination) during the COVID-19 pandemics. Then, an epidemic model tracks and projects case series Y_t based on current observations and other covariates such as vaccination awareness and acceptance (Figure 2). The functional response of these covariates can be mechanistic (i.e., parameters in SEIR-type model and rules in ABM) or data-driven, discussed below.

The first approach is to use this novel CSI to parameterize and calibrate mechanistic models, including SEIR-type compartment models and more recently introduced agent-based models (ABMs) that tracks detailed behaviors and interactions among individuals. We will compare and evaluate the relationship of content (CV_{tj}) and sentiment (SV_{tj}) time series with traditionally measured health outcomes, such as numbers of reported cases, hospitalizations, and deaths due to COVID-19. By parameterizing vaccination acceptance on these actual health outcomes, it will significantly enhance ABM's ability to further incorporate dynamic behavioral aspects, evaluate effectiveness of vaccination for COVID-19, and predict unintended consequences such as varying vaccination uptake rates across time and location.

Another major category of epidemic modeling is non-mechanistic data-driven models. Our previous study, along with several other studies, have shown that multivariate deep learning models, such as different types of recurrent neural network (RNN) models, can effectively project epidemic dynamics of COVID-19 (18, 19). Depending on different hypotheses, content (C) and sentiment (S) of interventions can be regarded as input variables that influence observed disease outcomes (D), such that $D_{tj} = f(C_{tj}, S_{tj})$. Alternatively, we could hypothesize no *a priori* influence, i.e., observed disease outcomes and online contents, sentiments toward interventions (e.g., vaccination) can mutually influence each other. Changing health outcomes in different phases of the pandemic can also influence public perceptions of the severity of the COVID-19 pandemic, and consequently alter vaccination acceptance. In this circumstance, multiple time series D_{tj}, C_{tj}, and

S_{tj} are modeled in parallel in RNN to make projections of each time series into the future.

3. Discussion

In this paper, we propose a more granular and comprehensive CSI as a critical component in the integrated disease surveillance system through effective data mining on online discourse data during an epidemic such as COVID-19. Social media and other online platforms provide massive data for knowledge discovery through advanced computational techniques, such as NLP. The dynamic changes in public awareness and perceptions toward various interventions, especially COVID-19 vaccination, can be effectively derived from NLP. Exploring these more granular dimensions of information, previously unavailable in traditional infoveillance, should significantly enhance integrative modeling efforts.

This proposed novel CSI framework naturally integrates theoretical foundations of social sciences and technical advances in information and computer science to address an important public health issue: to effectively incorporate cognitive and behavioral aspects into epidemic modeling. Here, we suggest some potential applications of the proposed infoveillance framework. It can effectively identify tipping points in public sentiments toward certain controversial topics, such as vaccination especially in the U.S. Knowing exactly when, where, and how the public will respond to COVID-19 vaccination can be crucial to inform local and national public health agencies to develop health communication strategies to encourage mass immunization and minimize the consequences of preventable cases, hospitalizations, and deaths. In addition, the novel CSI framework can be applied in conjunction with NLP-based misinformation detection methods to monitor surges of vaccination-related misinformation. This CSI framework could also evaluate responses and perceptions of different populations (e.g., race/ethnicity, age, or other social determinants of health) to specific types of interventions.

While social media provide large volumes of public discourse data on diseases to characterize public responses, sampling bias may still occur due to the observational study nature of passive infoveillance. Users of online platforms such as social media may not be adequately representative of the target population. Therefore, active participatory studies, such as randomized surveys, can complement this novel CSI *via* social media analytics.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This study was partially supported by the Models of Infectious Disease Agent Study (MIDAS) Network award MIDASUP-05.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proceed. Royal Soc. A.* (1927) 115:700–21. doi: 10.1007/bf02464423

2. Chen S, Robinson P, Janies D, Dulin M. Four challenges associated with current mathematical modeling paradigm of infectious diseases and call for a shift. *Open Forum Infect Dis.* (2020) 7:ofaa333. doi: 10.1093/ofid/ofaa333

3. The National Notifiable Disease Surveillance System. Centers for Disease Control and Prevention (CDC). Available online at: https://www.cdc.gov/nndss/index.html (accessed March 03, 2023).

4. COVID Data Tracker. CDC. Available online at: https://covid.cdc.gov/covid-data-tracker/#datatracker-home (accessed March 03, 2023).

5. Eysenbach G. Infodemiology and infoveillance: Framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. *J Med Int Res.* (2009) 11:1157. doi: 10.2196/jmir.1157

6. Eysenbach G. Infodemiology and infoveillance: tracking online health information and cyberbehavior for public health. *Am J Prevent Med.* (2011) 40(SUPPL) 2. doi: 10.1016/j.amepre.2011.02.006

7. Badell-Grau RA, Cuff JP, Kelly BP. Investigating the prevalence of reactive online searching in the COVID-19 pandemic: infoveillance study. *J Med Int Res.* (2020) 22:19791. doi: 10.2196/19791

8. Barros JM, Duggan J, Rebholz-Schuhmann D.The application of internet-based sources for public health surveillance (infoveillance): systematic review. *J Med Int Res.* (2020) 22:3680. doi: 10.2196/13680

9. Daughton AR et al. Mining and validating social media data for COVID-19-related human behaviors between January and July 2020: infodemiology study. *J Med Int Res.* (2021) 23:7059. doi: 10.2196/27059

10. Guy S, Ratzki-Leewing A, Bahati R. Social media: a systematic review to understand the evidence and application in infodemiology. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering. (2012) 91:1. doi: 10.1007/978-3-642-29262-0_1

11. Tang L, Bie B, Park SE, Zhi D. Social media and outbreaks of emerging infectious diseases: a systematic review of literature. *Am J Infect Cont*. (2018) 46:10. doi: 10.1016/j.ajic.2018.02.010

12. Huang X, Li Z, Jiang Y, Li X, Porter D. Twitter reveals human mobility dynamics during the COVID-19 pandemic. *PLoS ONE*. (2020) 15:e0241957. doi: 10.1371/journal.pone.0241957

13. Chen S, Zhou L, Song Y, et al. A novel machine learning framework for comparison of viral COVID-19-related Sina Weibo and Twitter posts: workflow development and content analysis. *J Med Internet Res.* (2021) 23:e24889. doi: 10.2196/24889

14. Miller M, Banerjee T, Muppalla R. What are people Tweeting about Zika? An exploratory study concerning its symptoms, treatment, transmission, and prevention. *JMIR Public Health Surveillance*. (2017) 3:e38. doi: 10.2196/publichealth.7157

15. Safarnejad L, Xu Q, Ge Y. Identifying influential factors in the discussion dynamics of emerging health issues on social media: computational study. *JMIR Public Health and Surveillance*. (2020) 6:7175. doi: 10.2196/17175

16. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, trends, and sentiments of tweets about the COVID-19 pandemic: temporal infoveillance study. *J Med Int Res*. (2020) 22:2624. doi: 10.2196/22624

17. Karafillakis E. Methods for social media monitoring related to vaccination: systematic scoping review. *JMIR Public Health and Surveillance*. (2021) 7:7149. doi: 10.2196/17149

18. Shahid F, Zameer A, Muneeb M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos Solitons Fractals*. (2020) 140:110212. doi: 10.1016/j.chaos.2020.110212

19. Chen S, Paul R, Janies D, Murphy K, Feng T, Thill JC, et al. Exploring feasibility of multivariate deep learning models in predicting COVID-19 epidemic. *Frontiers Public Health*. (2021) 9:661615. doi: 10.3389/fpubh.2021.661615

20. Lwin M, Lu J, Sheldenkar A, Schulz P, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on twitter: analysis of twitter trends. *JMIR Public Health Surveill*. (2020) 6:e19447. doi: 10.2196/19447

21. Geronikolou S, Drosatos G, Chrousos G. Emotional analysis of twitter posts during the first phase of the COVID-19 pandemic in Greece: infoveillance study. *JMIR Form Res*. (2021) 5:e27741.

22. Eckman P. An argument for basic emotions. *Cogn Emot*. (1999) 6:169–200. doi: 10.1080/02699939208411068

23. Kort B, Reilly R, Picard RW. An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion. *Proceed IEEE Int Conf Adv Learn Technol*. (2001) 3:43–6. doi: 10.1109/ICALT.2001.943850

frontiers | Frontiers in Public Health

TYPE Brief Research Report
PUBLISHED 23 March 2023
DOI 10.3389/fpubh.2023.1122230

OPEN ACCESS

EDITED BY
Pierpaolo Ferrante,
National Institute for Insurance against Accidents at Work (INAIL), Italy

REVIEWED BY
Xiangjun Du,
Sun Yat-sen University, China
Kayode Oshinubi,
Northern Arizona University, United States

*CORRESPONDENCE
Jorge P. Rodríguez
✉ jrodriguez@imedea.uib-csic.es;
✉ jorgeprodriguezg@gmail.com

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 12 December 2022
ACCEPTED 03 March 2023
PUBLISHED 23 March 2023

CITATION
Rodríguez JP, Aleta A and Moreno Y (2023) Digital cities and the spread of COVID-19: Characterizing the impact of non-pharmaceutical interventions in five cities in Spain. *Front. Public Health* 11:1122230. doi: 10.3389/fpubh.2023.1122230

Digital cities and the spread of COVID-19: Characterizing the impact of non-pharmaceutical interventions in five cities in Spain

Jorge P. Rodríguez[1,2]*, Alberto Aleta[2,3] and Yamir Moreno[2,3,4]

[1]Instituto Mediterráneo de Estudios Avanzados (IMEDEA), CSIC-UIB, Esporles, Spain, [2]Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, Zaragoza, Spain, [3]Department of Theoretical Physics, University of Zaragoza, Zaragoza, Spain, [4]CENTAI Institute, Turin, Italy

Mathematical modeling has been fundamental to achieving near real-time accurate forecasts of the spread of COVID-19. Similarly, the design of non-pharmaceutical interventions has played a key role in the application of policies to contain the spread. However, there is less work done regarding quantitative approaches to characterize the impact of each intervention, which can greatly vary depending on the culture, region, and specific circumstances of the population under consideration. In this work, we develop a high-resolution, data-driven agent-based model of the spread of COVID-19 among the population in five Spanish cities. These populations synthesize multiple data sources that summarize the main interaction environments leading to potential contacts. We simulate the spreading of COVID-19 in these cities and study the effect of several non-pharmaceutical interventions. We illustrate the potential of our approach through a case study and derive the impact of the most relevant interventions through scenarios where they are suppressed. Our framework constitutes a first tool to simulate different intervention scenarios for decision-making.

KEYWORDS

epidemic spreading, digital twins, COVID-19, non-pharmaceutical interventions, pandemic control

1. Introduction

The COVID-19 pandemic has globally impacted a plethora of systems, with health (1), socio-economic (2, 3), and environmental (4) consequences. To control the spread of SARS-CoV-2, policymakers implemented a diversity of procedures, grouped into either mitigation or suppression strategies. Lockdowns, implying home confinement, were frequently introduced to stop the spreading in early 2020 when the dynamics of the infection mechanisms were not clear. However, these lockdowns resulted in deep impacts on the economy, and later on, other non-pharmaceutical interventions were designed, such as the use of face masks, the closure of restaurants, universities, or schools, as well as contact tracing, testing, and isolation of close contacts of infected individuals.

The initial stages of the pandemic represented a high degree of uncertainty, both regarding the original transmission of the pathogen to human beings and reliable surveillance data [due to low testing efforts and inappropriate surveillance systems (5)]. Nowadays the situation has improved, as the availability of more data—even if many times of poor quality and low reliability—in principle allows to characterize the spreading at a large scale. Moreover, the existent data enables the development of mathematical models that help

quantify the observed evolution of the pandemic and evaluate the effects of the intervention scenarios.

The first wave of COVID-19 raised a challenge for modeling approaches due to the general bad data quality. Specifically, the lack of knowledge about the COVID-19 spread, the similarity between the symptoms of COVID-19 and those of influenza, and the low testing effort led together to lower rates of diagnosis and hence underreporting mainly in the number of cases (6), but also in the number of deaths. Seroprevalence studies (7) and the analysis of anomalies on the temporal series of deaths (8) were needed to estimate the real impact of the spreading process, showing that there were up to 10 times more cases than the reported ones. In this regard, spreading models can shed light on the real outcome of the infection across the population.

To properly model the spreading of a disease in the population, it is fundamental to acknowledge that human interactions are highly heterogeneous. Although network epidemiology can capture part of this diversity, such as the broad nature of the distribution of the number of interactions, the variability of contexts remains out of this formalism. These contexts can be effectively captured using multilayer networks, which are networks with multiple layers, each one describing the interactions in a different context (9, 10). In this work, we leverage anonymous, publicly available data to build high-resolution synthetic cities and encode them in multilayer networks (11, 12). We use these synthetic networks to study the propagation of the first wave of COVID-19 in five Spanish cities. Furthermore, we extend the simulation to the second wave for the particular case of the city of Zaragoza and thoroughly characterize the impact of non-pharmaceutical interventions during this period.

2. Materials and methods

2.1. Multilayer contact networks

We create five digital populations describing the inhabitants and the interactions between them in the cities of Barcelona, Valencia, Seville, Zaragoza, and Murcia, all of them located in Spain (Figures 1A–F). Their population ranges between 450 thousand and 1.7 million inhabitants (Figures 1G–I). Additionally, we include external individuals that may not be registered in the census but with most of their interactions expected to happen in these cities. These external individuals include old people living in nursing homes and non-local university students. Each inhabitant is represented in the population as a node connected to other inhabitants. These links were built according to the specific data sources for each city and each feature, as listed in the Supplementary material.

2.1.1. Demography

We obtained the geographical distribution, sex, and age of the inhabitants of the cities at the beginning of 2020 from multiple demographic data sources. The maximum spatial resolution was the census district (Figures 1B–F), at which we found most of the needed information to create the synthetic digital cities. Ages were available in age groups with a resolution of 5 years. Thus, we interpolated these age groups to consider a resolution of 1 year between 0 and 30 years, which was necessary to properly infer the interactions at schools and universities. This allowed us to create a synthetic population for each city resembling the characteristics of the real ones.

2.1.2. Contact networks

We modeled the contacts between individuals through networks described by the aggregation of multiple interaction layers (13). Specifically, we considered six interaction layers: home, nursing homes, school, work, university, and community. We incorporated empirical data from multiple datasets available from national, regional, and local sources to infer the connections between the individuals in the different environments introduced by the interaction layers. In Figure 2, we show the age mixing patterns of the population extracted from our synthetic cities (12).

2.1.3. Home layer

Individuals in the home layer are connected if they live together. We extracted the information on the number of homes of a specific size, the average home size, and the home structure at the district level. We use the information from the national census of 2011 (14), the most recent one that is currently available, (see also the Supplementary material) for all the cities except for Barcelona, for which this information is available from local sources with higher resolution. We also use the age difference of the home nuclei, at district resolution, from the national census. This information is key for reproducing realistic home contact matrices, as the home structures include in the "adults" category any individual aged between 25 and 64 years old. Connecting randomly pairs of individuals in this broad group could lead to less representative links in most homes, composed of two adults alone or with children or old people. As we do not know if these nuclei are assortative or disassortative, we include the data on the age difference of the nuclei to create this synthetic layer.

2.1.4. School layer

This layer connects all the students and a teacher within the same scholar unit. Besides, all the teachers that work in the same center are also connected. We included in this layer the infant levels (0–3 and 3–6 years old), primary school (7–12), secondary school (13–16), high school (17–18), and job training (from 17). We inferred these connections using data on the number of students per level, the number of units per level, and the number of schools, taking into account the levels offered by each kind of school. This information was available at the district level for Barcelona and Valencia. Additionally, data on the specific size of each specific unit at each center was available for Valencia and used for that layer inference. For Seville, Zaragoza, and Murcia, this information was available at the municipality level, so we mapped the school coordinates to the districts, and we inferred the rest of the needed information from the one at the municipality level. Once the synthetic units, at each center, were created, we assigned individuals from the population to those units. First, we filled each unit with individuals of that specific age that have their homes located in the same district. Secondly, the units that had not been filled totally with individuals from the same

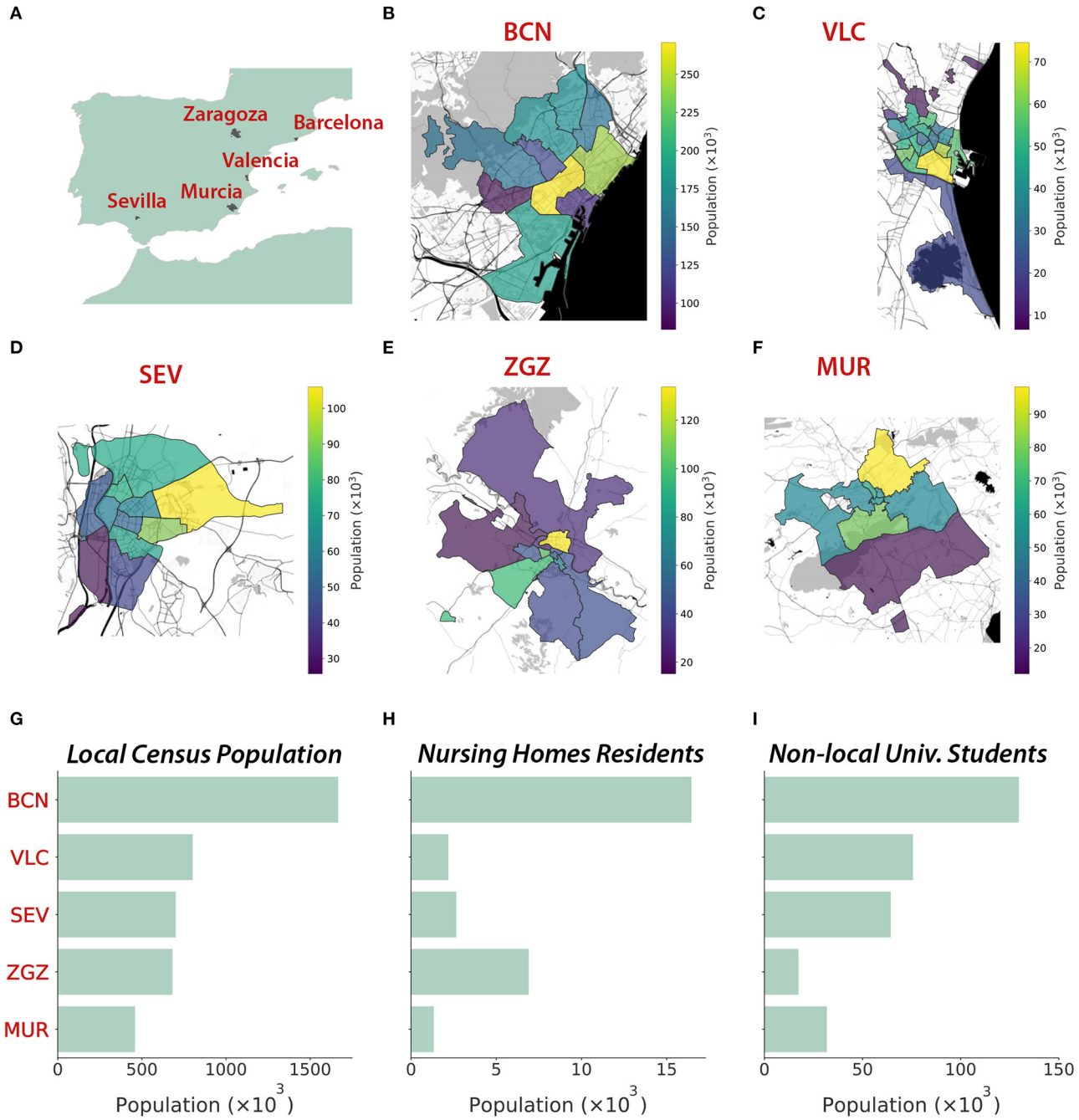


FIGURE 1
Geography and demography of the five cities represented through synthetic populations. **(A)** Location of the five cities. **(B–F)** Geographical representation (Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL) of the district structure at each city (source: Instituto Nacional de Estadística, INE), with colors representing the population of each district. **(G)** Local and non-local populations (nursing homes residents and university students). The five cities studied are Barcelona (BCN), Valencia (VLC), Seville (SEV), Zaragoza (ZGZ), and Murcia (MUR).

district were filled with individuals from other districts, with a priority determined by the distance between the centroids of the districts, until all the units in the city were full. We assumed that, after that step, the remaining individuals were not included in the education system. Teachers were chosen randomly among individuals aged between 30 and 70 from any district in the city.

2.1.5. University layer

We generated the university contact layer using the national statistics of the number of registered students per university and per degree, split by sexes, available from the Spanish Ministry of Education, considering both undergraduate and graduate programs, for the academic year 2019–2020. We considered the universities located in the same province as the studied

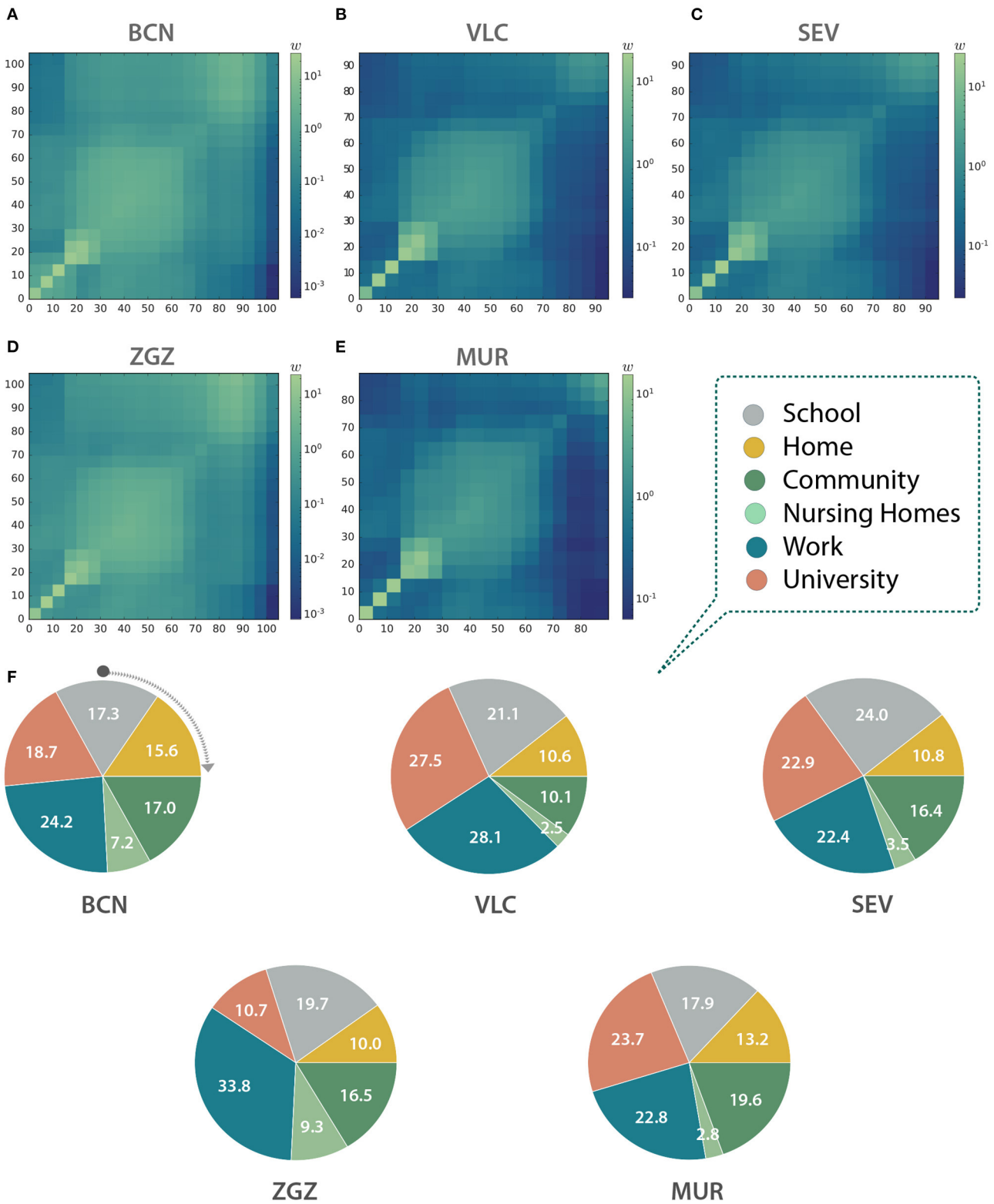


FIGURE 2
Contact matrices between age groups in **(A)** Barcelona (BCN), **(B)** Valencia (VLC), **(C)** Seville (SEV), **(D)** Zaragoza (ZGZ), and **(E)** Murcia (MUR). **(A–E)** Each entry w_{ij} is computed as the total number of observed links between individuals belonging to the age group in row i and column j, normalized with the number of individuals in each column. Thus, they represent, for a given column, the expected number of links of a random individual to individuals from each row. **(F)** Link distribution among the different layers in the five studied cities expressed as a percentage of the total number of contacts in each city.

cities, after removing distance-learning universities. Then, we estimated which students are registered and live in the city, and which ones are external, either registered in the same or another province. Local students have also interactions in the other layers, while externals are only in the university layer. Finally, we obtained the national age profiles by sex of university students, picked individuals from our synthetic population, and introduced those external, according to these profiles. We designed a connectivity pattern of all-to-all for degrees that had sizes lower than 50 people, and otherwise generated patches with all-to-all connectivity of a maximum size of 50 people for the larger ones.

2.1.6. Work layer

In the work layer, individuals that belong to the same company are connected together. We obtained the distribution of companies' sizes (S) throughout all the Spanish provinces, which follows a power-law distribution pdf $(S) \sim S^{-2}$. Then, we generated companies with sizes that follow this distribution, and, when sizes were higher than 20 people, we distributed the workers among patches with a maximum size of 20 people. We estimated the number of workers by subtracting the number of autonomous workers from the number of registered workers in each city, according to the Social Security reports. We extracted sex and age features also from Social Security reports on a national scale. We did not consider as potential workers those that were assigned a school patch, either as teachers or students. The synthetic companies were filled with individuals from the synthetic population following the corresponding distributions by age and gender.

2.1.7. Nursing homes layer

We collected information on the number of nursing homes and their capacity in each municipality. Additionally, we gathered national statistics on the age and gender of the people that reside in nursing homes. We assumed that the nursing homes need one caretaker for every four places, and chose that uniformly from those in the dataset older than 16 years old (minimum age for being allowed to work). Inside each nursing home, we assumed an all-to-all connection. Note that individuals residing in nursing homes do not interact in the household layer.

2.1.8. Community layer

We generated a synthetic community layer connecting randomly pairs of individuals living in the same district, according to the contact matrices for Spain in Prem et al. (15). There were contact matrices available for home, work, school, and other locations, and we chose the latter. This dataset reported the probability of connecting pairs of individuals according to their ages, in age groups of 5 years up to 75 years old. For individuals older than 75 years old, we extrapolated the data of the oldest available group.

2.2. Epidemic spreading

2.2.1. Spreading model

We used the COVASIM software for modeling the spread of COVID-19 (16). COVASIM is an open-source Python-based agent-based modeling tool. COVASIM considers a susceptible-exposed-infected-recovered or dead (SEIRD) epidemic model that includes disease parameters informed by the medical literature. The infected compartment is divided into asymptomatic and symptomatic infectious individuals, with the latter including presymptomatic, mild, severe, and critical stages. The three symptomatic stages can evolve to the recovered state, while the critical state can alternatively lead to the death of the individual (Supplementary Figure S1). The probabilities of developing symptoms, severe symptoms, a critical case, and from it the death of the individual are specified by age groups, arranged in 10-year-long age cohorts. This software has been used for studying different scenarios, for example, assessing the test-trace-quarantine strategy (17) or quantifying the risk of outbreaks after international border opening (18). We modified COVASIM to include the specific details of our synthetic cities. Specifically, we included the age, sex, and contacts of the individuals in each of the considered cities. We ran independent simulations where each simulation chose one randomly infected seed as the first infected individual. Then, we kept the endemic realizations, defined as those leading to a finite number of deaths, which we set higher than 10 for the first wave. For the second wave, we also requested that there were more than 10 death observations in the last 10 days of the realization. Apart from the internal parameters of COVASIM, we considered independently the infection rate and the date of arrival of the first infected individual. We calibrated both parameters for each city analyzing the official time series of deaths (see section 2.2.2), which were more reliable than the number of cases that could suffer from high underdetection rates (6).

2.2.2. Epidemic data

We obtained the temporal series of the number of deaths and the number of confirmed cases from the Spanish Ministry of Health (19) at the province level, with daily resolution. Then, we multiplied these values by the fraction of the province's population living in the city. We averaged the rescaled data over a moving window of 7 days (the specific day $\pm$ 3 days) to smooth the fluctuations.

2.3. Quantifying public health policies

To illustrate the potential of our approach, after calibrating the first wave of COVID-19 in our model with synthetic digital cities, we implemented a case study of the second wave focusing on the city of Zaragoza. This second wave occurred between July and December 2020.

The non-pharmaceutical interventions that were introduced in this city to mitigate the spread of COVID-19 were the following:

- Testing and tracing. Positive tested individuals and their close contacts were isolated for 14 days until 30th September, and from then on for 10 days.

- Restrictions on restaurants, cafes, and nightlife. Starting on 5th August, lifted on 4th September, and re-started on 19th October.
- Opening of the schools with reduced group sizes and safety measures. Different levels started progressively, from 7th September to 17th September.
- Opening of the universities with reduced group sizes. The university was opened on 14th September.
- Interventions impacting the community layer. We considered the interventions related to the State of Alarm and those related to the capacity and schedules of restaurants and bars.

In addition to these policies, we also took into account the annual leave period of workers and considered a reduced number of interactions in the work layer starting on 15th July until the middle of September, with the maximum reduction on 1st August. In terms of the model, this implies a higher amount of time (e.g., more weight of this layer) associated with the interactions in the community layer.

We used the infection rate (β) and date of arrival of the infected seed obtained in the calibration of the first wave and ran the model from the estimated arrival of the first case (see section 3.2) to 1st December 2020. Since we ran the simulations from the beginning of the first wave, we lifted progressively the restrictions that were active from the 14th of March in two steps: on the 1st and 20th of June, following the progressive lift of the restrictions that actually happened.

We kept 25% of the contacts in the school and university layers to simulate the small group's policy, and reduced β by 50% in the school layer, considering the strict protocol to avoid contagion at schools.

After calibrating the model of the second wave with these non-pharmaceutical interventions and considering the results of the simulation, we introduced different counterfactual scenarios in which we quantified the effect of each of the non-pharmaceutical interventions adopted. More specifically, we computed the number of deaths and the disease prevalence by performing simulations with the same epidemiological parameters but switching on and off alternative interventions. Finally, we computed the relative change in the relevant quantity X (X = deaths or prevalence) as $r = (X^{\text{counterfactual}} - X^{\text{simulated-2nd-wave}})/X^{\text{simulated-2nd-wave}}$. Therefore, the absolute change can be obtained as $X^{\text{counterfactual}} = (1 + r) \cdot X^{\text{simulated-2nd-wave}}$.

We considered the following nine different counterfactuals:

- No testing and no contact tracing. The testing intervention was removed. Hence, as contact tracing depends on the results of the testing process, contact tracing was automatically removed.
- No contact tracing. To analyze the impact of the contact tracing strategy and decouple it from the testing process, we kept the testing intervention and its parameters but removed contact tracing interventions.
- Opening 100% university. We considered the opening of the university layer with 100% of the contacts, instead of the 25% contacts estimated through the small group's intervention.
- Not opening university. We simulated a scenario where the university layer remained closed.
- Opening all schools together (x2). The school opening was done following a staggered strategy, such that each level started on a different date. We simulated scenarios where all the levels started on the same date, either on the date of the earliest opening (7th September) or the latest opening (17th September).
- 100% β in schools, whole groups. Schools were one of the sectors where strong protocols were introduced, reducing considerably the infection rate and also the group size. We simulated the absence of these protocols, keeping the same infection rate as in the rest of the layers, and considering this layer with whole groups, that is, 100% of the contacts.
- Not opening schools. We quantified the changes in the outcome of the second wave if the schools had not been open.
- No interventions in October. We observed that the interventions in October were key to controlling the second wave. Thus, we removed these interventions and computed this counterfactual, keeping the same final date, such that the result was comparable with the rest of the counterfactuals. However, we assumed that removing these interventions would imply that the second wave continued growing on time. To characterize this growth (in terms of both time extent and outcome), we ran additional simulations for 15 and 30 days more and compared them with extrapolations of the original second wave simulation for the same period (without including additional measurements introduced on December 2020 or calibrating the observed data in that period).

3. Results

3.1. Contact matrices

With the information contained in the multilayer networks we can infer the contact matrices of the population (12). These matrices can be used to inform classical epidemiological models for studies not based on agent-based models, or to obtain an aggregated picture of the interactions in the system, as in this case. Indeed, as we can see in Figures 2A–E, the shape of the matrices indicates that our networks display an assortative pattern with blocks of infants, adults, and the older adults with a higher preference to interact with other individuals with similar ages. The number of contacts per layer is also significantly different both within and across cities (Figure 2F). For instance, workplace contacts are predominant in Barcelona and Zaragoza, the university ones in Valencia and Murcia, and the school contacts in Seville. Note that our agent-based model explicitly contains each link between two individuals, and thus these matrices are not used to model the spreading.

3.2. First wave

In order to be able to explore realistic counterfactuals for the effectiveness of the most important NPIs adopted, we started by simulating the first wave to calibrate the model for each of the cities considered. Specifically, we ran simulations of the spread of COVID-19 in these cities using the software COVASIM. We estimated the transmission rate and the arrival of the initial seed,

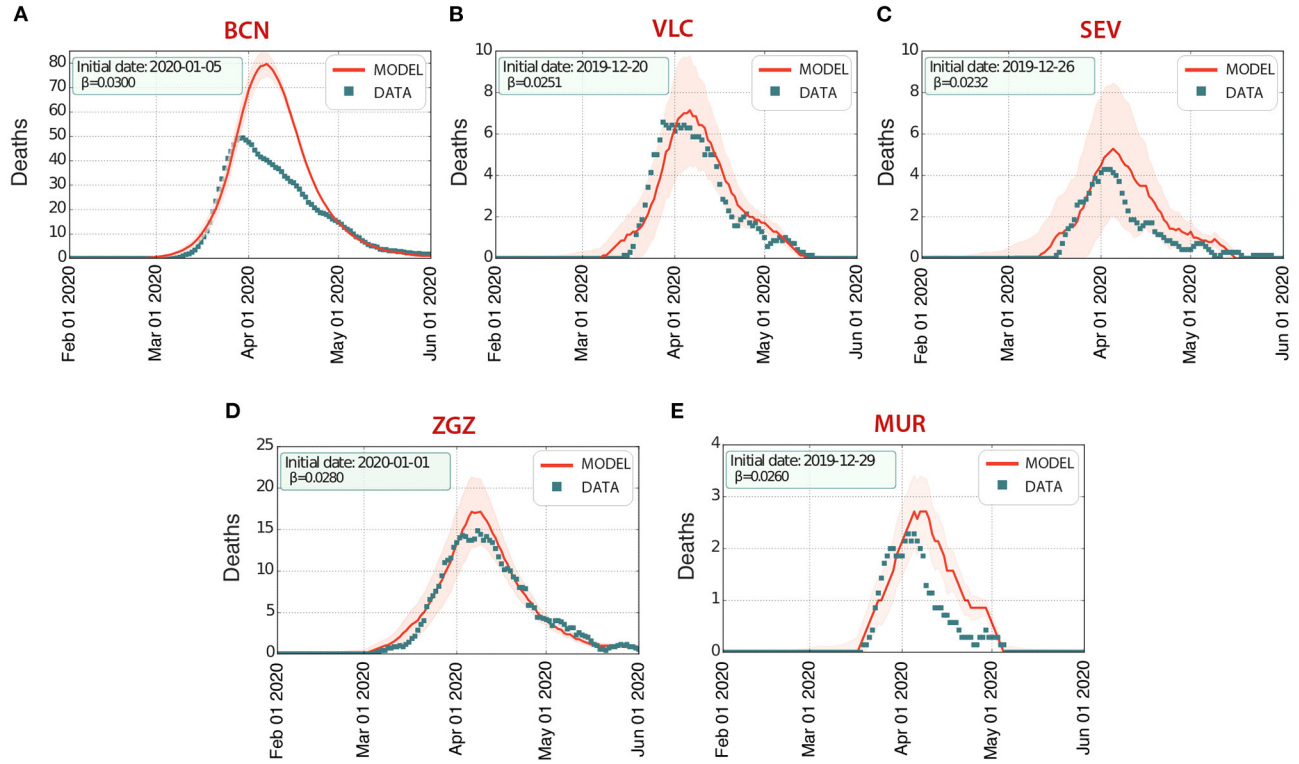


FIGURE 3
First wave of infections of COVID-19 in five Spanish cities: **(A)** Barcelona (BCN), **(B)** Valencia (VLC), **(C)** Seville (SEV), **(D)** Zaragoza (ZGZ), and **(E)** Murcia (MUR). The number of deaths *D* estimated by our model (solid line, the shaded area represents the 5-95% CI) agrees with the data (properly rescaled, see section 2) reported by the Spanish Ministry of Health (dots), with an excess of deaths in the initial stages of the wave.

TABLE 1 Starting date (Start), infection rate (β), prevalence (Prev.), and number of deaths (Deaths) from the model output of the first wave.

City	Start	β	Prev. (5–95% CI)	Deaths (5–95% CI)	Ref. Prev. (5–95% CI)
Barcelona	Jan 05	0.0300	14.6 (13.1–16.1)	2,334 (2,188–2,477)	7.4 (6.2–8.9)
Valencia	Dec 20	0.0251	3.1 (1.5–4.7)	227 (141–313)	2.1 (1.5–3.0)
Seville	Dec 26	0.0232	2.4 (1.1–4.7)	195 (75–314)	2.7 (1.9–3.8)
Zaragoza	Jan 1	0.0280	6.1 (3.9–8.3)	599 (465–733)	5.2 (3.9–6.9)
Murcia	Dec 29	0.0260	2.4 (1.6–3.3)	89 (67–111)	1.6 (1.0–2.5)

The reference prevalence (Ref. Prev.) is that provided in the first phase of the national study of seroprevalence for the first wave in its second round, finished on June 1st (7).

considered as a single infected individual (Figure 3, Table 1). Our multilayer approach allowed us to introduce the effects of the national lockdown declared on 14th March 2020, reducing the contacts in the work layer to 20% (10% for Barcelona) and 0% in the university, school, and community layers. Our results indicate an earlier arrival of COVID-19 to these cities (upon the assumption of a single initial seed), and they highlight the earlier occurrence of deaths at the beginning of the first wave, not considered in the official statistics, in Barcelona, Valencia, Seville, and Zaragoza. The prevalence estimates from our model are compatible with those obtained from the nationwide seroprevalence study in Spain (see Table 1). Note that this seroprevalence study detected 10 times more cases than the ones reported by the surveillance system.

3.3. Second wave. Counterfactuals

When restrictions were progressively lifted after the end of the first wave, a second wave started growing (Figure 4), and we calibrated our model to obtain the impact of the interventions on our model parameters (see section 2).

The calibration of the second wave led to the following results:

- Varying number of links in the community layer: connections were set at 50% of the baseline value (1st June, progressive lift of restrictions), 80% (20th June, end of the national State of Alarm), 50% (5th August, regional limits on restaurant and bar schedules), 100% (4th September, lift of restrictions), 30% (19th October, regional limits on the schedule and capacity of

restaurants and bars), and 10% (26th October, national State of Alarm).

- Varying the weight of the links in the community layer: increase by 50% from 20th June to 26th October (end and beginning, respectively, of the national State of Alarm).
- Varying the number of links in the work layer: connections were set at 70% of the baseline value (1st June), 50% (15th July), 30% (1st August), and 50% (20th September). These changes in summer accounted for the summer holidays period. Mobility reports (20) showed a slow return to mobility associated with work in September.
- Testing. The probability of symptomatic individuals being tested (per day with symptoms) was estimated to be 15% between 1st July and 14th September, and 9% from 15th September. The delay between the test and the result notification (with the beginning of the associated isolation period) was fixed to 1 day.
- Contact tracing. The contacts from positive-tested individuals were traced with a general probability P_t. Additionally, P_t was weighted for each layer of contacts, fixing the weights 1 for home, 0.8 for school, 0.6 for university, 0.8 for work, 0.0 for the community [until the introduction of the contact tracing app Radar COVID (21), which increased it to 0.05], and 0.0 for nursing homes. We fixed the time between the positive notification and the communication with close contacts to 2 days. P_t was estimated to be 0.4 between 1st July and 19th August, 0.45 between 20th August (introduction of Radar COVID) and 30th September, and 0.5 from 1st October (extra support to contact tracing from trained soldiers).

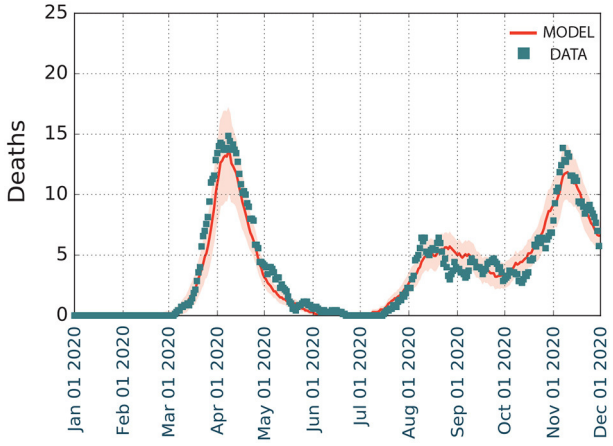


FIGURE 4
Modeling the first and second waves of COVID-19 spreading on Zaragoza from January to December 2020. We represent the temporal evolution of the number of deaths D, with shaded regions depicting the 5–95% confidence intervals of the model.

Overall, the model (Figure 4) estimated that there were 1,354 deaths (5–95 CI), with a prevalence of 22.6% (21.0–24.2% 5–95 CI). This prevalence was higher than the reported in the fourth phase of the national seroprevalence study carried out in mid November (7), which estimated a prevalence of 12.7% (10.1–15.8% 5–95 CI) at the province level. Even though the data at the municipality level is not available, it reported that the prevalence in municipalities with more than 100,000 inhabitants was 50% larger than in the smaller ones. The province of Zaragoza is highly heterogeneous in terms of size, with one municipality (out of 293) containing 69% of the almost 1,000,000 inhabitants in the province. Thus, it is expected that the prevalence at the city level should be much larger. Similarly, our model in the first wave agreed with the empirical observations of the temporal evolution of the number of deaths documented, with a minor overestimation for Murcia but a larger one for Barcelona. We interpret these divergences as possibly missing data, in line with other studies that have claimed a higher number of deaths than that reported by the official statistics, which was particularly significant in the administrative region of Catalonia, where Barcelona is located (8).

The results of the counterfactual analysis shown in Figure 5 indicate that the combination of tracing and testing, with the associated isolation of positive individuals, was very effective in reducing the number of both deaths and infections. Note that for this case, the counterfactual (i.e., lack of such measures) led to more than twice the number of infected individuals, and also nearly doubled the number of deaths. Next, we quantified the impact of contact tracing alone by keeping the testing process, together with the isolation of individuals with a positive test, but removing the contact tracing. This scenario also showed an increase in both the number of infections and deaths. However, the increase of both observables was twice lower than if both contact-tracing and testing are removed, highlighting the importance of combining these two interventions to achieve the best result. The third-most important counterfactual according to the increase in deaths was the removal of the interventions in schools, which however produce the second-largest increase in the number of infections, but with a smaller impact in the number of deaths because most infections would occur in non-risk age groups. Interestingly, opening the university without restrictions would lead to fewer infections than with schools completely opened, however leading to more deaths. The synchronous opening of the schools for different levels also implied an increase in prevalence and deaths, but with a lower impact than other interventions. Finally, there were also some counterfactual scenarios that produced a decrease in the observables, such as keeping schools or the university closed. However, their impact was minor.

For the sake of completeness, we also assessed the impact of the interventions in October to control the outbreak. In principle, these measures did not have a big impact as shown in Figure 5. However, the interpretation is not straightforward, because our second wave simulations finish on December 1st, and the absence of these interventions may have implied a later end of this wave. To quantify this, we extended, without adding any new intervention, both the calibrated and counterfactual simulations, finishing the simulation on (a) 15th December and (b) 31st December. Our results showed that the relative change between the real extrapolated framework and this counterfactual kept increasing after 1st December (9.5% for deaths, 50.9% for prevalence), as the extrapolated values were higher on 15th December (40.3% for deaths, 71.1% for prevalence), and slightly decreased at 31st December (25.3% for deaths, 66.5%

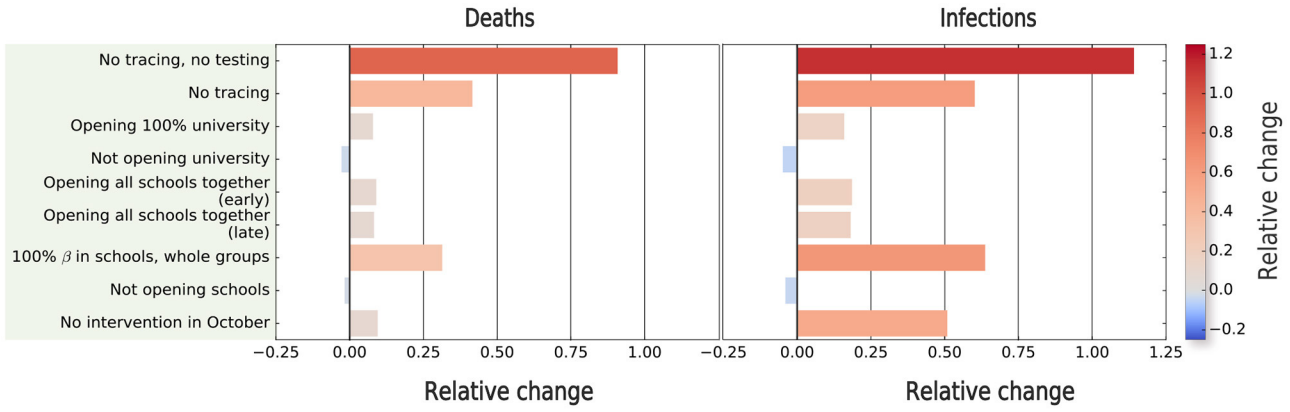


FIGURE 5
Estimating the impact of non-pharmaceutical interventions on the second wave of COVID-19 spreading in Zaragoza. We compute the relative change with respect to the outcome of the model in terms of prevalence and number of deaths for different simulated scenarios. The relative difference in the relevant quantity X (X = deaths or prevalence) is computed as $(X^{\text{counterfactual}} - X^{\text{simulated-2nd-wave}})/X^{\text{simulated-2nd-wave}}$.

for prevalence), indicating that by this date, the counterfactual wave would have finished. Hence, this extrapolation suggests that this counterfactual would have implied, compared with the rest of the counterfactuals, the second-largest increase in prevalence, and the fourth-largest increase in the number of deaths.

4. Discussion and conclusions

The quick spread of a deadly infection among a population represents a threat to public health systems, which requires immediate action and extra resources in order to mitigate and eventually control the impact of the associated disease on the population. However, interventions need to be carefully evaluated, as our society is a complex interdependent system in which mitigating the effects of one disease might result in non-desirable side effects such as the reduction of the services provided to both prevent and treat other diseases (22–24). To avoid these effects, scientists and policy-makers need computational resources that allow for a fast assessment of the possible outcome of interventions whenever pharmaceutical interventions, such as vaccination campaigns, are not possible. This is often the case when new diseases emerge, as seen with the virus SARS-CoV-2, which currently represents one of the major threats to public health systems. On the one hand, governments and health organizations need to allocate significant resources in order to develop and test vaccines and/or specific pharmacological treatments. On the other hand, traditional surveillance methods are likely to miss large numbers of new infections in the population, leading to a high number of undocumented infections during the early stages of the disease, as also happened in the case of SARS-CoV-2 (25). In such situations, non-pharmaceutical interventions represent the alternative to at least earning time. This is where data-driven and computational frameworks are fundamental to inform models that can illustrate the outcome of different non-pharmaceutical interventions. In this paper, we have presented a model that could be used to characterize the consequences of a plethora of non-pharmaceutical interventions. We applied the model to study the first and second waves of COVID-19 in Spain, finding that testing, tracing, and isolation were among the most effective interventions to reduce both the number of deaths and infections, in line with similar studies for other geographical locations (13, 26, 27). Our study also shows that, on the whole, the interventions adopted during the second wave for the city of Zaragoza, were effective and reduced the number of deaths and infections by around 10% and 50%, respectively. The effort presented in this work, informing a computational model of COVID-19 spreading with synthetic populations based on real data, has the potential to speed up the analysis of different intervention scenarios in future large-scale epidemic emergencies.

This work has some limitations that deserve further discussion. First of all, our simulations considered a single randomly chosen initial seed, and from this, we estimated the date of arrival of the disease. Nevertheless, the spreading process could have started by the arrival of several infected individuals either synchronously or asynchronously. However, we think that these approaches are equivalent, as they would lead to the same number of infected individuals at later dates. In contrast, different effects could emerge when specific individuals, according to, for example, their age, district of residence, or employment status, display a higher likelihood to introduce the infection. Another limitation is the isolation of the cities, as they are considered closed systems. This can be solved by introducing a spontaneous infection rate reflecting the imported cases from other locations, although we assume that in the cases of generalized local transmission this rate would lead to minor differences. The emergence of variants with different infection and recovery rates and death probabilities is challenging for these models, requesting the parameter correction for subsequent waves happening when other variants were present. The latter, however, does not impact predictions at the early stages of an emerging disease, which is when evaluating possible NPIs is most needed. Finally, we considered the main non-pharmaceutical interventions that were applied in the city of Zaragoza, but their calibration may include the effect of other interventions that we assumed to have minor effects.

Our multilayer network method, informed from multiple data sources, contrasts with other approaches used for modeling the spread of COVID-19. These include the introduction of meta-population approaches representing recurrent mobility (28, 29), the use of demography to infer social contact data (6), information from real-time human mobility indices (30), or the use of high-resolution individual trajectories (13). We acknowledge that the latter method would be the ideal scenario in terms of accuracy, but it would request the availability of detailed mobility data, which is not directly linked to layers whose dynamics are shaped by interventions. When such data is not available, our method can inform mathematical models of spreading while keeping realistic social contact data.

In summary, our work shows how models of digital cities can be coupled to agent-based epidemiological models of disease dynamics and be used for scenario evaluation. Our approach aligns with the spirit of developing digital twins to face the challenges raised by the Sustainable Development Goals (https://sdgs.un.org), related for example with environmental problems or health issues. After the extensive data search needed for creating these cities (see Supplementary material), updating these digital cities will be a more straightforward task, allowing them to timely inform the models that help design non-pharmaceutical interventions to mitigate the effects of future pandemics.

Data availability statement

The data sources are specified in the Supplementary material. The code used for simulating the first and second waves in Zaragoza is available in the following URL: https://github.com/jorgeprodriguezg/digicovid. Digital Cities are available from JR (jorgeprodriguezg@gmail.com) upon reasonable request.

Author contributions

JR, AA, and YM contributed to conception and design of the study and analyzed the results. JR collected, cleaned, organized the data, performed the statistical analysis, and wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

JR, AA, and YM acknowledge support from the Government of Aragon (FONDO–COVID19-UZ-164255). JR is supported by Juan de la Cierva Formacion program (Ref. FJC2019-040622-I) funded by MCIN/AEI/ 10.13039/501100011033. AA acknowledges support through the grant RYC2021-033226-I funded by MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR. YM was partially supported by the Government of Aragon, Spain and ERDF A way of making Europe through grant E36-20R (FENOL), and by Ministerio de Ciencia e Innovación, Agencia Española de Investigación (MCIN/AEI/10.13039/501100011033) Grant No. PID2020-115800GB-I00. The authors acknowledge the use of the computational resources of COSNET Lab at Institute BIFI, funded by Banco Santander through grant Santander-UZ 2020/0274, and by the Government of Aragon (FONDO–COVID19-UZ-164255). JR and AA acknowledge funding from la Caixa Foundation under the project code SR20-00386 (COVID-SHINE). The funders had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

The authors acknowledge the Department of University Statistics of the Spanish Ministry of Universities for facilitating the number of registered students at the university per municipality, which is not currently publicly available online.

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.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1122230/full#supplementary-material

References

1. Barber RM, Sorensen RJ, Pigott DM, Bisignano C, Carter A, Amlag JO, et al. Estimating global, regional, and national daily and cumulative infections with SARS-CoV-2 through Nov 14, 2021: a statistical analysis. *Lancet.* (2022) 399:2351–80. doi: 10.1016/S0140-6736(22)00484-6

2. Bonaccorsi G, Pierri F, Cinelli M, Flori A, Galeazzi A, Porcelli F, et al. Economic and social consequences of human mobility restrictions under COVID-19. *Proc Nat Acad Sci USA.* (2020) 117:15530–5. doi: 10.1073/pnas.2007658117

3. Pak A, Adegboye OA, Adekunle AI, Rahman KM, McBryde ES, Eisen DP. Economic consequences of the COVID-19 outbreak: the need for epidemic preparedness. *Front Publ Health.* (2020) 8:241. doi: 10.3389/fpubh.2020.00241

4. Bates AE, Primack RB, Biggar BS, Bird TJ, Clinton ME, Command RJ, et al. Global COVID-19 lockdown highlights humans as both threats and custodians of the environment. *Biol Conserv.* (2021) 263:109175. doi: 10.1016/j.biocon.2021.109175

5. Starnini M, Aleta A, Tizzoni M, Moreno Y. Impact of data accuracy on the evaluation of COVID-19 mitigation policies. *Data Policy.* (2021) 33:e28. doi: 10.1017/dap.2021.25

6. Pullano G, Di Domenico L, Sabbatini CE, Valdano E, Turbelin C, Debin M, et al. Underdetection of cases of COVID-19 in France threatens epidemic control. *Nature.* (2021) 590:134–9. doi: 10.1038/s41586-020-03095-6

7. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Hernán MA, Pérez-Olmeda M, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet.* (2020) 396:535–44. doi: 10.1016/S0140-6736(20)31483-5

8. García-García D, Vigo MI, Fonfría ES, Herrador Z, Navarro M, Bordehore C. Retrospective methodology to estimate daily infections from deaths (REMEDID) in COVID-19: the Spain case study. *Sci Rep.* (2021) 11:11274. doi: 10.1038/s41598-021-90051-7

9. Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA. Multilayer networks. *J Compl Netw.* (2014) 2:203–71. doi: 10.1093/comnet/cnu016

10. Aleta A, Moreno Y. Multilayer networks in a nutshell. *Annu Rev Condens Matter Phys.* (2019) 10:45–62. doi: 10.1146/annurev-conmatphys-031218-013259

11. Fumanelli L, Ajelli M, Manfredi P, Vespignani A, Merler S. Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread. *PLoS Comput Biol.* (2012) 8:e1002673. doi: 10.1371/journal.pcbi.1002673

12. Mistry D, Litvinova M, Pastore y Piontti A, Chinazzi M, Fumanelli L, Gomes MFC, et al. Inferring high-resolution human mixing patterns for disease modeling. *Nat Commun.* (2021) 12:323. doi: 10.1038/s41467-020-20544-y

13. Aleta A, Martin-Corral D, Pastore y Piontti A, Ajelli M, Litvinova M, Chinazzi M, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat Hum Behav.* (2020) 4:964–71. doi: 10.1038/s41562-020-0931-9

14. Censo de Población y Viviendas, Instituto Nacional de Estadística. (2011). Available online at: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176992&menu=ultiDatos&idp=1254735576757#:~:text=Censos%20de%20Poblaci%C3%B3n%20y%20Viviendas%202011.,y%20alcanza%20los%204.193.319 (accessed August 4, 2021).

15. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput Biol.* (2017) 13:e1005697. doi: 10.1371/journal.pcbi.1005697

16. Kerr CC, Stuart RM, Mistry D, Abeysuriya RG, Rosenfeld K, Hart GR, et al. Covasim: an agent-based model of COVID-19 dynamics and interventions. *PLOS Comput Biol.* (2021) 17:e1009149. doi: 10.1371/journal.pcbi.1009149

17. Kerr CC, Mistry D, Stuart RM, Rosenfeld K, Hart GR, Núñez RC, et al. Controlling COVID-19 via test-trace-quarantine. *Nat Commun.* (2021) 12:2993. doi: 10.1038/s41467-021-23276-9

18. Pham QD, Stuart RM, Nguyen TV, Luong QC, Tran QD, Pham TQ, et al. Estimating and mitigating the risk of COVID-19 epidemic rebound associated with reopening of international borders in Vietnam: a modelling study. *Lancet Glob Health.* (2021) 9:e916–24. doi: 10.1016/S2214-109X(21)00103-0

19. Red Nacional de Vigilancia Epidemiologica (Spain) (2022). Available online at: https://cnecovid.isciii.es/covid19 (accessed March 16, 2021).

20. Google. *Google COVID-19 Community Mobility Reports* (2021). Available online at: http://www.google.com/covid19/mobility (accessed April 26, 2021).

21. Rodríguez P, Graña S, Alvarez-León EE, Battaglini M, Darias FJ, Hernán MA, et al. A population-based controlled experiment assessing the epidemiological impact of digital contact tracing. *Nat Commun.* (2021) 12:587. doi: 10.1038/s41467-020-20817-6

22. Roberts L. Pandemic brings mass vaccinations to a halt. *Science.* (2020) 368:116–7. doi: 10.1126/science.368.6487.116

23. Jewell BL, Mudimu E, Stover J, Ten Brink D, Phillips AN, Smith JA, et al. Potential effects of disruption to HIV programmes in sub-Saharan Africa caused by COVID-19: results from multiple mathematical models. *Lancet HIV.* (2020) 7:e629–40. doi: 10.1016/S2352-3018(20)30211-3

24. Tovar M, Aleta A, Sanz J, Moreno Y. Modeling the impact of COVID-19 on future tuberculosis burden. *Commun Med.* 2:77. doi: 10.1038/s43856-022-00145-0

25. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science.* (2020) 368:489–93. doi: 10.1126/science.abb3221

26. Steinbrook R. Contact tracing, testing, and control of COVID-19-learning from Taiwan. *JAMA Intern Med.* (2020) 180:1163–4. doi: 10.1001/jamainternmed.2020.2072

27. Salathé M, Althaus CL, Neher R, Stringhini S, Hodcroft E, Fellay J, et al. COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. *Swiss Med Wkly.* (2020) 150:w20225. doi: 10.4414/smw.2020.20225

28. Eguíluz VM, Fernández-Gracia J, Rodríguez JP, Pericàs JM, Melián C. Risk of secondary infection waves of COVID-19 in an insular region: the case of the Balearic Islands, Spain. *Front Med.* (2020) 7:563455. doi: 10.3389/fmed.2020.563455

29. Arenas A, Cota W, Gómez-Gardeñes J, Gómez S, Granell C, Matamalas JT, et al. Modeling the spatiotemporal epidemic spreading of COVID-19 and the impact of mobility and social distancing interventions. *Phys Rev X.* (2020) 10:041055. doi: 10.1103/PhysRevX.10.041055

30. Kraemer MU, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science.* (2020) 368:493–7. doi: 10.1126/science.abb4218

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 31 March 2023
DOI 10.3389/fpubh.2023.1111641

OPEN ACCESS

EDITED BY
Pierpaolo Ferrante,
National Institute for Insurance Against Accidents at Work (INAIL), Italy

REVIEWED BY
Arvind Ramanathan,
Argonne National Laboratory (DOE), United States
Youcef Mammeri,
Laboratoire Amiénois de Mathématique Fondamentale et Appliquée (LAMFA), France

*CORRESPONDENCE
Patricio Cumsille
✉ pcumsille@ubiobio.cl

SPECIALTY SECTION
This article was submitted to Infectious Diseases: Epidemiology and Prevention, a section of the journal Frontiers in Public Health

RECEIVED 29 November 2022
ACCEPTED 02 March 2023
PUBLISHED 31 March 2023

CITATION
Cumsille P, Rojas-Díaz O and Conca C (2023) A general modeling framework for quantitative tracking, accurate prediction of ICU, and assessing vaccination for COVID-19 in Chile. *Front. Public Health* 11:1111641. doi: 10.3389/fpubh.2023.1111641

A general modeling framework for quantitative tracking, accurate prediction of ICU, and assessing vaccination for COVID-19 in Chile

Patricio Cumsille[1,2]*, Oscar Rojas-Díaz[3] and Carlos Conca[2,4]

[1]Department of Basic Sciences, Faculty of Sciences, University of Bío-Bío, Chillán, Chile, [2]Centre for Biotechnology and Bioengineering, University of Chile, Santiago, Chile, [3]Department of Mathematics and Computers Science, Faculty of Science, University of Santiago of Chile, Santiago, Chile, [4]Department of Mathematical Engineering and Center for Mathematical Modeling, University of Chile (UMI CNRS 2807), Santiago, Chile

Background: One of the main lessons of the COVID-19 pandemic is that we must prepare to face another pandemic like it. Consequently, this article aims to develop a general framework consisting of epidemiological modeling and a practical identifiability approach to assess combined vaccination and non-pharmaceutical intervention (NPI) strategies for the dynamics of any transmissible disease.

Materials and methods: Epidemiological modeling of the present work relies on delay differential equations describing time variation and transitions between suitable compartments. The practical identifiability approach relies on parameter optimization, a parametric bootstrap technique, and data processing. We implemented a careful parameter optimization algorithm by searching for suitable initialization according to each processed dataset. In addition, we implemented a parametric bootstrap technique to accurately predict the ICU curve trend in the medium term and assess vaccination.

Results: We show the framework's calibration capabilities for several processed COVID-19 datasets of different regions of Chile. We found a unique range of parameters that works well for every dataset and provides overall numerical stability and convergence for parameter optimization. Consequently, the framework produces outstanding results concerning quantitative tracking of COVID-19 dynamics. In addition, it allows us to accurately predict the ICU curve trend in the medium term and assess vaccination. Finally, it is reproducible since we provide open-source codes that consider parameter initialization standardized for every dataset.

Conclusion: This work attempts to implement a holistic and general modeling framework for quantitative tracking of the dynamics of any transmissible disease, focusing on accurately predicting the ICU curve trend in the medium term and assessing vaccination. The scientific community could adapt it to evaluate the impact of combined vaccination and NPIs strategies for COVID-19 or any transmissible disease in any country and help visualize the potential effects of implemented plans by policymakers. In future work, we want to improve the computational cost of the parametric bootstrap technique or use another more efficient technique. The aim would be to reconstruct epidemiological curves to predict the combined NPIs and vaccination policies' impact on the ICU curve trend in real-time, providing scientific evidence to help anticipate policymakers' decisions.

KEYWORDS

COVID-19, predictive modeling, epidemiological modeling, time delays, vaccination, practical identifiability, parameter optimization, parametric bootstrap

1. Introduction

The COVID-19 pandemic has induced a significant research effort for tracking, prediction, and control. In Chile, which is no stranger to the above, health authorities initiated vaccination in the summer of 2021, gradually reducing overall ICU patients and death by COVID-19 while suspending non-pharmaceutical interventions (NPIs) such as lockdowns (partial or total). One of the main lessons is that we must be prepared to face another pandemic like it. Consequently, this article aims to develop a general modeling framework consisting of epidemiological modeling generalization and devising a practical identifiability approach to assess combined vaccination and NPIs strategies for the dynamics of any transmissible disease. To validate the framework, we applied it to track COVID-19 dynamics, accurately predict the ICU curve trend in the medium term, and assess vaccination in Chile.

The literature on COVID-19 modeling is vast. A search in the Web of Science (WOS) with the terms "COVID-19", "modeling", and "time delays", refined by WOS categories related to STEM disciplines, resulted in 95 articles (in January 2023). Therefore, we only review some of those that we believe are important for their applications. For example, Al-Tuwairqi and Al-Harbi (1) proposed a model to investigate the effects of time delay in vaccine production on COVID-19 spread. In addition, Zhenzhen et al. (2) studied a model with "long memory" to describe the multi-wave peaks of the COVID-19 dynamics, where "long memory" allows for predicting this last using non-local terms, which means that one can include an arbitrary long history of the disease. Indeed, for a particular non-local term, the authors obtained a model with time delays. Furthermore, the authors modeled vaccination as an impulsive term that translates into decreased susceptibility.

Moreover, Ghosh et al. (3) derived a model with time delay, where the last is the disease duration, i.e., the average time in which infected individuals recover or die. Zhai et al. (4) investigated a SEIR-type model with time delay and vaccination control. The first parameter is similar to that introduced in our previous work (5) but is considered in the exposed population equation. They simulated vaccination as a control that decreases susceptibility, similar to the generalization we propose in the present work. Finally, our current work relies on generalizing the model developed by (5), which has common elements with some cited works here. Indeed, in (5), we introduced the same time delay that models the average time to recover or die, as in (3). At the same time, we could interpret our previous model (5) as one incorporating a long memory effect in the sense that it allows the reproduction of multi-wave peaks depending on the parameter values, as we showed.

The general goal is to implement a hybrid approach (6), in this case, a holistic combination of mathematical modeling with a practical identifiability approach to reconstruct and predict epidemiological curves based on careful optimization, synthetic data, automatic data scanning, and calibration. Precisely, the scientific novelty of the article relies on developing a general modeling framework that could contribute to anticipating epidemiological scenarios, evaluating the impact of combined vaccination and NPI strategies for any transmissible disease, and helping to visualize the potential effects of implemented plans by policymakers.

To achieve the general goal, we rely on our previous work that forecasted COVID-19's second wave in May 2021 in Chile, calibrating data between March and September 2020 (before vaccination began) through suitable epidemiological modeling (5). Our specific goals are:

1. To generalize our previously developed epidemiological modeling to describe vaccination and assess combined vaccination and NPI strategies for the dynamics of any transmissible disease and, as a study case, of COVID-19.
2. To improve the practical identifiability approach to calibrate the generalized epidemiological modeling with different datasets representing different regions of Chile and accurately forecast the ICU curve trend in the medium term in any stage of the COVID-19 pandemic.
3. To provide open-source codes that implement our general epidemiological modeling framework with standardized parameter initialization for every dataset for reproducibility.

To implement the general modeling framework, we processed and used the official COVID-19 datasets provided by the Chilean government. The framework consists of two main parts: epidemiological modeling generalization and devising a practical identifiability approach. The first consists of a non-linear delay differential equations (DDE) system describing time variation and transitions between the compartments of susceptible, infected, recovered and the sum of ICU plus dead. The second relies on parameter optimization (5, 7, 8), a parametric bootstrap technique (9), and data processing. A novelty of this work is the implementation of a careful parameter optimization algorithm by searching for suitable initializations according to each processed dataset. In addition, we implemented a parametric bootstrap technique to accurately predict the ICU curve trend in the medium term and assess vaccination.

We have organized the article as follows: we describe the general modeling framework in Section 2. More precisely, based on our previous work (5), we describe the epidemiological modeling generalization in Section 2.1. Then, in Section 2.2, we detail the careful parameter optimization algorithm implementation, the parametric bootstrap technique, and data processing, among other methodology pieces. Then, we provide the modeling results to validate the framework with several datasets representing COVID-19 dynamics in different regions of Chile in Section 3. Finally, in Section 4, we discuss the results and give some conclusions in Section 5.

2. General modeling framework' description

In this section, we describe the general modeling framework. To do so, we have split it into two main subsections that generalize previously developed epidemiological modeling and devise the practical identifiability approach.

2.1. Epidemiological modeling generalization

In the present section, we describe the epidemiological modeling introduced by (5), then present and provide a complete description of each part of the modeling generalization.

2.1.1. Previous work

To derive the epidemiological modeling generalization, we rely on the *generalized SIR model with constant time delays* or *generalized SIR model* previously devised by (5), which describes the NPIs' effect through variations in the rate of disease transmission. We remark here that the NPIs' impact consists of social distancing and that the generalized SIR model does not describe vaccination. To be precise, the model corresponds to the following non-linear DDE system:

$$\frac{dS}{dt}(t) = -\frac{\beta(t)}{N}S(t)I(t-\tau_1), \tag{1a}$$

$$\frac{dI}{dt}(t) = \frac{\beta(t)}{N}S(t)I(t-\tau_1) - \gamma_{IR}I(t-\tau_2), \tag{1b}$$

$$\frac{dR}{dt}(t) = \gamma_{IR}I(t-\tau_2). \tag{1c}$$

It is worth noting that the generalized SIR model can generate complex dynamics since, by contrast to the classical SIR model, it can simulate more than one local maximum for the infected. Then, it provides a way to explain several COVID-19 waves, depending on the parameters' values (5). In addition, the generalized SIR model would produce better prediction results than the classical SEIR model since no observation of the exposed population is available since, similarly to the asymptomatic population, those exposed are challenging to observe. Indeed, the Chilean government's COVID-19 database (10), apart from the symptomatic cases, counts the asymptomatic ones, and there is no way to know how many of these become symptomatic (5).

The parameters of the model (1) are as follows: $\beta(t)$ corresponds to the mean rate of disease transmission, γ_{IR} is the mean removal rate, τ_1 is the mean incubation time of disease, and τ_2 is the mean time from onset to clinical recovery or death caused by disease, or the duration time of disease until recovery or death.

Model (1) follows the susceptible-infectious-removed (SIR) paradigm. Susceptible (S) individuals infected by SARS-CoV-2 undergo incubation during a mean time (τ_1) before becoming infected (I). The infected individuals are infected by the disease for a mean time (τ_2) until being removed (R) by clinical recovery or death.

The initial conditions have to satisfy $S(t_0) + I(t_0) + R(t_0) = N$, where N is the size of the population under study for a closed system and taking into account that $(S + I + R)'(t) = 0$ for all $t > 0$, for a suitably chosen t_0.

In the case of COVID-19, following the discussion by (5), we assume that the number of infected reported with symptoms confirmed by Reverse Transcriptional Polymerase Chain Reaction (RT-PCR) tests, denoted by $I_r(t)$, is underestimated since it depends on the availability and application of RT-PCR tests. Consequently, we assume that $I_r(t)$ is a fraction of the *actual number of infected* $I(t)$,

$$I_r(t) = f(t)I(t), \tag{2}$$

where $f(t)$ is the ratio of positive RT-PCR tests number (confirmed cases) over the actual infected cases for the day t, which accounts for the *real positivity rate*. It is worth noting that $f(t)$ is not the same as the positivity rate of detected cases, but it is related to it, and it involves the asymptomatic infected (5). We modeled $f(t)$ as an inverted Sigmoid-type function such that if $I(t)$ is small enough, which occurred during the beginning of the outbreak, then an important fraction of the real infected cases are detected ($I_r(t) \approx I(t)$). On the contrary, when $I(t)$ is large enough, which occurred just before the quarantines were imposed, then only a small fraction $0 < a < 1$ of the real infected cases are detected ($I_r(t) \approx aI(t)$). Precisely, $f(t)$ is defined as

$$f(t) = 1 + \frac{a-1}{1 + e^{-k(I(t)-I_{thr})}} \tag{3}$$

whose parameters are a, k and I_{thr}, where a and k represent the minimum and the decay rate of $f(t)$, respectively. On the other hand, to measure how large/small $I(t)$ is, we introduce a threshold I_{thr} such that $I(t) \ll I_{thr}$ implies $f(t) \approx 1$, and $I(t) \gg I_{thr}$ implies $f(t) \approx a \ll 1$.

Model (1), describing time variation and transitions S-to-I-to-R plus Equations (2)–(3), representing the real positivity rate, is quite general since it models the dynamics of any transmissible disease, not considering vaccination. The model parameters, gathered in vector $(\beta(t), \gamma_{IR}, \tau_1, \tau_2, a, k, I_{thr})$, are unknown or inaccessible and have to be identified from time-series observations I_r corresponding to the number of infected reported with symptoms confirmed by RT-PCR tests. In (5), we devised a practical identifiability approach, i.e., a set of techniques to reliably estimate parameters with acceptable accuracy from noisy data (11, 12). In particular, we reproduced epidemiological scenarios considering $\beta(t)$ varying in time to describe NPI strategies ranging from total relaxation to imposing strict social distancing (complete lockdown differed by municipalities) at different periods. As a result, we forecasted the second wave of May 2021 in Chile, calibrating data between March and September 2020 (before vaccination began).

2.1.2. Toward modeling's generalization

To make model (1) more complete and realistic, we developed a generalization to measure the hospital load on the healthcare system through the number of patients hospitalized in the ICU and to assess the vaccination. Equations (2)–(3) that model the real positivity rate remain unchanged. In the following, we provide a modeling description of both aspects.

2.1.2.1. Vaccination description

According to (13), vaccination protects in four ways: against infection, symptoms, severe disease, and reducing onward transmission. However, even considering part of the vaccination's ways of protection, the model can become very complex, as in the work by (13). Moreover, since COVID-19 data has great uncertainty, among other issues discussed in the data processing section, any model will provide results accordingly, no matter how exact its representation of reality is. Consequently, our present

work is in the same spirit as our previous work (5) in that keeping a model simple is critical to reasonably carrying out a practical identifiability approach. Even so, it is still a challenge to do it in real-time to anticipate epidemiological scenarios to help predict the hospital load on the healthcare system, as required in the first year of the pandemic (before vaccination). In this regard, we assume vaccination protects against infection by diminishing susceptibility, as assumed, for example, by (2, 4), which translates into adding new terms to Equations (1a) and (1c) for those susceptible and those recovered and several meaningful parameters associated with vaccination.

According to the previous discussion, the new terms that model vaccination account for the transition from susceptible to recovered, its inverse, and their respective times of transition. Transition rates are denoted by $\gamma_{SR}(t)$ and $\gamma_{RS}(t)$, while the times model as delays within the new terms added, designated by τ_3 and τ_4, respectively. We allow transition rates to vary in time to describe them more realistically since they depend on several factors that may change over time, as explained below (e.g., the immunity of an individual without booster doses decays faster).

Parameter $\gamma_{SR}(t)$ relates mainly to the vaccination uptake rate, while $\gamma_{RS}(t)$ relates to the waning of immunity after vaccination. Moreover, parameters τ_3 and τ_4 depend on every delivered vaccine's effectiveness and immunity waning. However, the vaccination uptake rate, effectiveness, and immunity waning of vaccines are not well-determined since they depend on several factors such as prior infection, age, sex, T-cell response, and the periodicity of vaccine injections. In addition, only natural infection mounts a significant and lasting immune response (14). Therefore, parameters $\gamma_{SR}(t)$, $\gamma_{RS}(t)$, τ_3, and τ_4 depend on many factors, which makes it difficult to estimate them and they are strongly available data-dependent, independent of how exact the model's representation of reality is. Thus, they are inaccessible and must be identified from time-series observations of COVID-19, as noted before.

The model equations that consider vaccination are:

$$\frac{dS}{dt}(t) = -\frac{\beta(t)}{N}S(t)I(t-\tau_1) - \gamma_{SR}(t)S(t-\tau_3) + \gamma_{RS}(t)S(t-\tau_4), \tag{4a}$$

$$\frac{dR}{dt}(t) = \gamma_{IR}(t)I(t-\tau_2) + \gamma_{SR}(t)S(t-\tau_3) - \gamma_{RS}(t)S(t-\tau_4) + \gamma_{UR}(t)U(t-\tau_6). \tag{4b}$$

The second and third terms of Equations (4a)–(4b) model the transitions from the susceptible to recovered compartment, and conversely, τ_3 stands for the mean time delay for those susceptible to become immune after vaccination, and $\gamma_{SR}(t)$ indicates how fast it happens. Similarly, τ_4 designates the mean time until an individual loses immunity, so τ_4 is the mean duration of immunity by vaccination, and $\gamma_{RS}(t)$ measures how fast it happens. Finally, the last term in Equation (4b) pertains to the U compartment, which we explain below.

2.1.2.2. *U* compartment description and modeling generalization summary

Finally, as mentioned before, we added the variable U to model (1), representing the sum of the patients in the ICU plus those confirmed dead due to COVID-19, the equation of which contains the transitions from I-to-U and U-to-R. The variable R now describes the recovered, whereas R in the model (1) represented the removed, i.e., the sum of those who had recovered plus those who had died.

As before, the mentioned transitions encompass rates and time delays, denoted by $\gamma_{IU}(t)$, τ_5, $\gamma_{UR}(t)$, and τ_6. Precisely, we model the transition I-to-U by the rate $\gamma_{IU}(t)$ and the time τ_5 that those infected took to be admitted to the ICU. In addition, we represent the transition U-to-R by the rate $\gamma_{UR}(t)$ and the time τ_6 that patients took to recover in the ICU.

We summarize the modeling generalization as the following non-linear DDE system:

$$\frac{dS}{dt}(t) = -\frac{\beta(t)}{N}S(t)I(t-\tau_1) - \gamma_{SR}(t)S(t-\tau_3) + \gamma_{RS}(t)S(t-\tau_4), \tag{5a}$$

$$\frac{dI}{dt}(t) = \frac{\beta(t)}{N}S(t)I(t-\tau_1) - \gamma_{IR}(t)I(t-\tau_2) - \gamma_{IU}(t)I(t-\tau_5), \tag{5b}$$

$$\frac{dR}{dt}(t) = \gamma_{IR}(t)I(t-\tau_2) + \gamma_{SR}(t)S(t-\tau_3) - \gamma_{RS}(t)S(t-\tau_4) + \gamma_{UR}(t)U(t-\tau_6), \tag{5c}$$

$$\frac{dU}{dt}(t) = \gamma_{IU}(t)I(t-\tau_5) - \gamma_{UR}(t)U(t-\tau_6). \tag{5d}$$

Again, the previous system is closed since the variables' sum equals N, the size of the targeted population. The DDE system (5) and Equations (2)–(3) will be named *general epidemiological modeling*, which is quite broad since it models the dynamics of any transmissible disease under any combination of NPIs and vaccination. To describe simply the in-time-variation of parameters $\beta(t)$, $\gamma_{IU}(t)$, and $\gamma_{UR}(t)$, we assumed that these are piecewise linear functions. Then, the functions $\beta(t)$, $\gamma_{IU}(t)$, and $\gamma_{UR}(t)$ are represented by the vectors $\boldsymbol{\beta}$, $\boldsymbol{\gamma_{IU}}$ and $\boldsymbol{\gamma_{UR}}$ in $\mathbb{R}^{n_\beta+1}$ that represent n_β straight lines approximating the respective functions. We gave the same description for the in-time-variation of parameters $\gamma_{SR}(t)$, $\gamma_{RS}(t)$, and $\gamma_{IR}(t)$, i.e., they are represented by the vectors $\boldsymbol{\gamma_{SR}}$, $\boldsymbol{\gamma_{RS}}$, and $\boldsymbol{\gamma_{IR}}$ in $\mathbb{R}^{n_\gamma+1}$.

Therefore, general epidemiological modeling depends on $p := 3(n_\beta + n_\gamma) + 15$ parameters gathered in the vector $\boldsymbol{\theta} \in \mathbb{R}^p$ defined by:

$$\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) \in \mathbb{R}^p, \tag{6a}$$

$$\boldsymbol{\theta_1} = (\boldsymbol{\gamma_{IR}}, \boldsymbol{\gamma_{SR}}, \boldsymbol{\gamma_{RS}}, \boldsymbol{\gamma_{IU}}, \boldsymbol{\beta}, \boldsymbol{\gamma_{UR}}, \boldsymbol{\tau})^t \in \mathbb{R}^{3(n_\gamma+n_\beta)+12}, \tag{6b}$$

$$\boldsymbol{\gamma_{IR}}, \quad \boldsymbol{\gamma_{SR}}, \quad \boldsymbol{\gamma_{RS}} \in \mathbb{R}^{n_\gamma+1}, \tag{6c}$$

$$\boldsymbol{\beta}, \quad \boldsymbol{\gamma_{IU}}, \quad \boldsymbol{\gamma_{UR}} \in \mathbb{R}^{n_\beta+1}, \tag{6d}$$

$$\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6)^t \in \mathbb{R}^6, \tag{6e}$$

$$\boldsymbol{\theta_2} = (a, k, I_{thr})^t \in \mathbb{R}^3, \tag{6f}$$

where $n_{\boldsymbol{\gamma}}$, $n_{\boldsymbol{\beta}}$ is the number of time intervals to reconstruct $\gamma_{SR}(t)$, $\gamma_{RS}(t)$, and $\gamma_{IR}(t)$, and $\beta(t)$, $\gamma_{IU}(t)$, $\gamma_{UR}(t)$ piecewise linearly with equally spaced intervals, respectively. For instance, using $n_{\boldsymbol{\beta}} = 9$, $n_{\boldsymbol{\gamma}} = 9$, one has $p = 3(n_{\boldsymbol{\beta}} + n_{\boldsymbol{\gamma}}) + 15 = 69$ parameters to estimate, i.e., $\boldsymbol{\theta} \in \mathbb{R}^{69}$.

Table 1 summarizes the parameters of general epidemiological modeling given by Equations (5), (2), and (3).

Table 1 Parameters of general epidemiological modeling.

Symbol	Description	Unit
τ_1	Mean incubation time	days
τ_2	Mean time to recover for mild cases	days
τ_3	Mean time from susceptible to recovery (by vaccination immunity)	days
τ_4	Mean duration of immunity (by vaccination)	days
τ_5	Mean time from infected to ICU	days
τ_6	Mean time from ICU to recover	days
γ_{IR}	Mean recovery rate for mild cases	days^{-1}
β	Mean transmission rate	days^{-1}
γ_{IU}	Mean transition rate from infected to ICU	days^{-1}
γ_{UR}	Mean recovery rate for ICU patients	days^{-1}
γ_{SR}	Mean transition rate from susceptible to recovered	days^{-1}
γ_{RS}	Mean transition rate from recovered to susceptible	days^{-1}
a	Minimum of the real positivity rate	–
k	Decay rate of the real positivity rate	inhabitants^{-1}
I_{thr}	Infection threshold of the real positivity rate	inhabitants

2.2. A practical identifiability approach

We devised a practical identifiability approach that relies on parameter optimization, a parametric bootstrap technique, and data processing, for which computer implementation includes open-source data and code repository through GitHub (15). The approach relies on solving a parameter estimation problem through a careful optimization algorithm, numerical resolution of modeling equations, a parametric bootstrap technique, and data processing. For solving the modeling equations, we required the provision of reasonable bounds for the parameters, which is critical for achieving a stable numerical method. In addition, the parameter range is meaningful, at least regarding the time delays that describe relevant parameters from the epidemiology viewpoint.

Next, we describe each piece of the practical identifiability approach.

2.2.1. Parameter estimation problem description

To reproduce and predict COVID-19 dynamics in Chile, one has to solve the *parameter estimation problem:* given a dataset of the time-series observations of COVID-19 dynamics, identify the parameter vector $\boldsymbol{\theta}$ such that general epidemiological modeling fits them in the least-squares sense. The time-series observations we used are

$$\left\{\left[(I_r)_j,\ (U_r)_j\right] : j = 1, \cdots, n\right\}.$$

I_r corresponds to the number of infected reported with symptoms confirmed by RT-PCR tests [see Equation (2) and its respective explanation], U_r corresponds to the observations of variable U, i.e., the sum of the patients in the ICU and confirmed deaths due to COVID-19, and n is the number of data.

More precisely, we have to find the vector $\boldsymbol{\theta} \in \mathbb{R}^p$, defined by Equation (6), that minimizes the sum of squares:

$$SS(\boldsymbol{\theta}) := \left\|(Res_I,\ Res_U)^t\right\|^2 = \sum_{j=1}^{n}\left[(Res_I)_j^2 + (Res_U)_j^2\right] \quad (7)$$

where Res_I and $Res_U \in \mathbb{R}^n$ stand for the relative residuals of variable I and U, respectively, defined by

$$(Res_I)_j := \frac{\left[(I_r)_j - f(t_j, \boldsymbol{\theta_2})I(t_j, \boldsymbol{\theta_1})\right]}{f(t_j, \boldsymbol{\theta_2})I(t_j, \boldsymbol{\theta_1})} \quad j = 1, \cdots, n \quad (8a)$$

$$(Res_U)_j := \frac{\left[(U_r)_j - U(t_j, \boldsymbol{\theta_1})\right]}{U(t_j, \boldsymbol{\theta_1})} \quad j = 1, \cdots, n. \quad (8b)$$

The objective function defined in (7) corresponds to the sum of squares of the residuals relative to the model observations, $f(t_j, \boldsymbol{\theta}_2)I(t_j, \boldsymbol{\theta}_1)$ and $U(t_j, \boldsymbol{\theta}_1)$. The choice of the relative residuals obeys to take into account the unequal quality of the observations (16). In this case, the patients in the ICU plus confirmed deaths due to COVID-19 (variable U) is better observed than those infected (variable I).

The components of Res_I, defined in (8a), correspond to the differences between the time-series observations $(I_r)_j$, and the model observations $f(t_j, \boldsymbol{\theta_2})I(t_j, \boldsymbol{\theta_1})$ (see Equation 2). The components of Res_U, defined in (8b), are the differences between the time-series observations $(U_r)_j$, and the model observations $U(t_j, \boldsymbol{\theta_1})$. Both variables, I and U, correspond to the solution of the general epidemiological model (5), (2), and (3) calculated at $(t_j, \boldsymbol{\theta})$ for $j = 1, \cdots, n$, for a given parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) \in \mathbb{R}^p$ defined by (6).

2.2.2. A careful optimization algorithm

To calibrate general epidemiological modeling, we implemented a careful optimization algorithm that combines data processing algorithms (cleaning, smoothing, and curve interpolation), initial parameter estimation generation, and a non-linear least-squares optimization method to minimize $SS(\boldsymbol{\theta})$ [Equation (7)] for each processed dataset.

The argument minimum of the sum of squares designs as $\hat{\boldsymbol{\theta}}$ for every dataset. The vector $\hat{\boldsymbol{\theta}}$ is called the *non-linear least-squares estimator*, abbreviated as *non-linear LSE*.

Next, we provide details on the implementation.

2.2.2.1. Implementation

We employed the *Trust-Region Interior Reflective* (TIR) method implemented in Matlab R2022b as the subroutine *lsqnonlin*, specially adapted for solving non-linear least-squares minimization problems. The convergence of the TIR method depends strongly on initial parameter estimation, which has to be relatively close to

the optimal solution (7, 8). We efficiently minimized the objective function by implementing a *percentage decrease technique* from the parameters' ranges to calculate suitable initial parameter vectors for every dataset in a standardized manner. According to our previous experience (5), the initial parameters that mainly influence the fitting results are $\boldsymbol{\beta}$, $\boldsymbol{\gamma_{IU}}$, $\boldsymbol{\gamma_{UR}} \in \mathbb{R}^{n_\beta+1}$, which describe the mean rate of disease transmission and transitions from I-to-U and U-to-R. To avoid overfitting, we selected n_β equispaced intervals to accurately fit and predict after the final calibration time for every dataset.

Next, we explain the numerical resolution of general epidemiological modeling within a range for meaningful parameters and the percentage decrease technique, which are critical for implementing our careful optimization algorithm.

2.2.2.2. Numerical resolution of general epidemiological modeling

To evaluate the objective function $SS(\boldsymbol{\theta})$, we numerically solved the model (5) at t_j, $j = 1, \cdots, n$, for different parameter vectors $\boldsymbol{\theta}$ chosen *ad-hoc* for each dataset. We carried out the numerical resolution by a Runge-Kutta type formula (17): the subroutine *dde23* implemented in Matlab, designed for solving non-linear DDE systems. In addition, we reconstructed the function of history (required for solving DDE instead of the initial condition for classical ordinary differential equations systems) for the model (5) by interpolating the data $I_r(t)$, $U_r(t)$, and recovered for every studied dataset. We used a shape-preserving piecewise cubic interpolation as devised by the *interp1* Matlab subroutine with the option *pchip*.

2.2.2.3. Meaningful parameter bounds

We computed the model parameters $\boldsymbol{\theta}$ given in (6) using meaningful bounds from an epidemiological viewpoint for the time delays: $1 \leq \tau_1 \leq 14$, $1 \leq \tau_2 \leq 21$, $14 \leq \tau_5 \leq 56$, and $21 \leq \tau_6 \leq 42$ days. We imposed these bounds because the incubation period (τ_1) ranges from 1 to 14 days (mean of 5–6 days), the median time from onset to clinical recovery for mild COVID-19 cases is approximately 2 weeks (τ_2), and is 3–6 weeks for patients with severe or critical symptoms (τ_6). In addition, among patients who died, the time from symptom onset to outcome ranged from 2 to 8 weeks (τ_5) (18). In addition, we assume that vaccination immunity duration, τ_4, ranges from 1 to 240 days since immunity declines only at 6–8 months after natural infection (19). In contrast, no range is well-determined for the transition time from susceptible to recovered, τ_3, however, one may expect that it is relatively small, so we assume $1 \leq \tau_3 \leq 14$.

For the real positivity fraction, $f(t)$, one has that

$$0 < a < 1, \quad \min\{(I_r)_j : j = 1, \cdots, n\} \leq I_{thr} \leq \max\{(I_r)_j : j = 1, \cdots, n\}.$$

Finally, all the rest of the parameters (transition and transmission rates) have to be within ranges to achieve stability of the careful optimization algorithm implementation, mainly related to general epidemiological modeling numerical resolution, as we explain next.

2.2.2.4. Percentage decrease technique for numerical stability

Convergence of the optimization algorithm, lsqnonlin, depends directly on that of the dde23 solver, both implemented in Matlab. Through our experiments, we verified that the model's numerical solution calculated by dde23 strongly depends on the derivatives of the first points evaluated, and its convergence relies on the closeness of the initial curves used to build the history function. To overcome this stability problem without intervening or designing new Matlab numerical libraries, we developed a simple but effective technique, called the *percentage decrease technique*, to produce initial curves contained in the feasible space of the official data curves.

The percentage decrease technique consists of multiplying the upper bounds of the parameter vector $\boldsymbol{\theta}$ (see Equation 6) by a fraction $\omega \in [1e-4, 1e-2]$, excluding time delays $\boldsymbol{\tau}$ (see Equation 6e). Concretely, we chose the initial parameter vector defined by $\boldsymbol{\theta}^{(0)} := \omega \cdot \mathrm{UB}_{\boldsymbol{\theta}}$, where $\mathrm{UB}_{\boldsymbol{\theta}}$ stands for the upper bound of $\boldsymbol{\theta}$, described in Section 2.2.2.3.

We calibrated ω for every studied dataset representing a characteristic epidemiological curve, with the magnitude or period of peaks' duration differentiated, which is mainly related to the density and mobility of the population. We chose a value of ω inversely proportional to the population size within the interval $[1e-4, 1e-2]$. Therefore, one should use a small value, e.g., $\omega = 1e-3$ or 0.1% of $\mathrm{UB}_{\boldsymbol{\theta}}$ for large cities (millions of inhabitants), and an even smaller value, e.g., $\omega = 1e-4$ or 0.01% of $\mathrm{UB}_{\boldsymbol{\theta}}$ for towns with fewer inhabitants (less than one million). Generally, $\omega = 1e-3$ works well for all cases obtaining a similar minimum error. Still, we calibrated a suitable value of ω for every dataset to speed up the execution time of the optimization subroutine lsqnonlin.

2.2.3. A parametric bootstrap technique

Once we calibrated a dataset, we applied a parametric bootstrap technique (PBT) to quantify the parameters' uncertainty and construct confidence intervals to achieve reliable and accurate forecasting performance, obtained by propagating the uncertainty (9). The PBT generates synthetic datasets repeatedly sampled from the least-squares curves:

$$F\left(t_j, \hat{\boldsymbol{\theta}}\right) := \left[f\left(t_j, \hat{\boldsymbol{\theta}}_2\right) I\left(t_j, \hat{\boldsymbol{\theta}}_1\right), \; U\left(t_j, \hat{\boldsymbol{\theta}}_1\right)\right]. \tag{9}$$

However, the PBT requires intensive computational resources and time since it generates simulated data from $F(t_j, \hat{\boldsymbol{\theta}})$ and calculates least-squares parameter estimates for each generated synthetic dataset. Therefore, we applied this technique to show the predictive power of general epidemiological modeling and assess vaccination only for one dataset, corresponding to days 280–530 that encompass the Metropolitan region's second wave, and considering the last 12 weeks to test the forecasting performance of the PBT.

In the PBT, one assumes that synthetic data follows a given probability distribution with an expected value equal to the least-squares curves $F(t_j, \hat{\boldsymbol{\theta}})$. To be precise, we implemented the following algorithm:

1. We calculated the parameter estimates $\hat{\boldsymbol{\theta}}$ through least-squares fitting the model to the time-series data to obtain the best-fit model given by $F(t_j, \hat{\boldsymbol{\theta}})$ (see Equation 9).
2. Using the least-squares fitted model $F(t_j, \hat{\boldsymbol{\theta}})$, we generated M replicated *synthetic datasets* denoted by $F_1^{SD}(t_j, \hat{\boldsymbol{\theta}})$, $F_2^{SD}(t_j, \hat{\boldsymbol{\theta}})$, $\cdots$, $F_M^{SD}(t_j, \hat{\boldsymbol{\theta}})$. We generated synthetic datasets as random vectors with a mean equal to $F(t_j, \hat{\boldsymbol{\theta}})$:

$$F_k^{SD}(t_j, \hat{\boldsymbol{\theta}}) \sim \mathbf{Dist}\left[F(t_j, \hat{\boldsymbol{\theta}})\right] \tag{10}$$

   where **Dist** is a given probability distribution of mean equal to $F(t_j, \hat{\boldsymbol{\theta}})$, and variance proportional to the mean magnitude or the covariance matrix of it. For instance, we used the normal, Poisson, and negative binomial distributions. The last is adequate to model data over-dispersion while controlling its magnitude (9).
3. We re-calculated the least-squares parameter estimates fitting the model to each of the M-simulated datasets realizations. Each parameter vector is denoted by $\hat{\boldsymbol{\theta}}_\ell$ for $\ell = 1, 2, \cdots, M$.
4. Using the set of re-estimated parameters $\hat{\boldsymbol{\theta}}_\ell$, $\ell = 1, 2, \cdots, M$, we calculated a confidence interval at the level of 95%. The uncertainty around the least-squares model fit is given by $F(t_j, \hat{\boldsymbol{\theta}}_1)$, $F(t_j, \hat{\boldsymbol{\theta}}_2), \cdots$, $F(t_j, \hat{\boldsymbol{\theta}}_M)$.

Typical values for the number of bootstrap samples M range from 50 to 200 for a proper standard error estimation; see p. 13–14 in (20). Indeed, by choosing $M = 100$, the standard error estimate provides reliable results for parameter estimation, as shown in Section 4. However, beyond choosing a good number of bootstrap samples M, careful implementation of the PBT is critical for obtaining $\hat{\boldsymbol{\theta}}_\ell$ corresponding to $F(t, \hat{\boldsymbol{\theta}}_\ell)$ so that the curves that are reproduced and predicted remain positive for all t within the time interval of calibration and prediction.

The calculated confidence interval is our prediction interval, denoted by PI, and defined by:

$$\begin{aligned} \mathrm{PI} :&= \left[\mathrm{LB}_{\hat{\boldsymbol{\theta}}},\ \mathrm{UB}_{\hat{\boldsymbol{\theta}}}\right] := \left[\bar{\hat{\boldsymbol{\theta}}} \pm t_{0.975,M-1}\ \mathbf{SE}\right] \\ &= \bar{\hat{\boldsymbol{\theta}}}\left[1 \pm t_{0.975,M-1}\ \mathbf{NSE}\right], \qquad (11a) \\ \mathbf{NSE} :&= \frac{\mathbf{SE}}{\bar{\hat{\boldsymbol{\theta}}}} = \frac{1}{\sqrt{M}}\frac{s_{\hat{\boldsymbol{\theta}}}}{\bar{\hat{\boldsymbol{\theta}}}}, \qquad (11b) \end{aligned}$$

where $\bar{\hat{\boldsymbol{\theta}}} \in \mathbb{R}^p$ is the mean of the M bootstrap parameters $\hat{\boldsymbol{\theta}}_\ell \in \mathbb{R}^p$, $t_{0.975,M-1}$ is the 0.975 percentile of the t-student distribution with $M-1$ degrees of freedom, $s_{\hat{\boldsymbol{\theta}}} \in \mathbb{R}^p$ is the standard deviation of the M bootstrap parameters, and $\mathbf{NSE} \in \mathbb{R}^p$ is the normalized standard error (5) (the definitions of $s_{\hat{\boldsymbol{\theta}}}$ and **NSE** are understood component-by-component).

The resulting uncertainty around the least-squares model fit, $F(t, \hat{\boldsymbol{\theta}})$, is quantified by the 95% confidence bounds [$\mathrm{LB}_{\hat{\boldsymbol{\theta}}}$ and $\mathrm{UB}_{\hat{\boldsymbol{\theta}}}$, defined by (11)] (9). Since we are interested in computing a more accurate prediction for the ICU curve trend in the medium term, we define an error criterion that includes performance for fit and forecasting. More precisely, we define

$$\begin{aligned} E(\hat{\boldsymbol{\theta}}_\ell) :&= 0.6\mathrm{FP}_U(\hat{\boldsymbol{\theta}}_\ell) + 0.2\mathrm{FP}_I(\hat{\boldsymbol{\theta}}_\ell) + 0.1[\mathrm{RMSE}_U(\hat{\boldsymbol{\theta}}_\ell) \\ &\quad + \mathrm{RMSE}_I(\hat{\boldsymbol{\theta}}_\ell)],\ \ell = 1, \cdots, M. \end{aligned} \tag{12}$$

In Equation (12), the root mean squared error (RMSE) to measure the fit performance is defined by (16):

$$\begin{aligned} \mathrm{RMSE}_I &= \left[\frac{1}{n-p}\sum_{j=1}^{n}(Res_I)_j^2\right]^{1/2}, \\ \mathrm{RMSE}_U &= \left[\frac{1}{n-p}\sum_{j=1}^{n}(Res_U)_j^2\right]^{1/2}. \end{aligned} \tag{13}$$

We calculated the RMSE over the calibration period (n and p are the calibrated dataset size and the number of model parameters, respectively). A similar criterion was employed for the forecasting performance (FP), but the sum over the prediction period was computed and normalized by the number of predicted data points.

In Equation (12), we gave more weight (60%) to FP_U since variable U is better observed than variable I, followed by FP_I (20%) and the RMSE for both variables (10% each). Furthermore, it was more important to make more effort to follow up with the sum of the ICU patients and those who had died than those who were infected. Thus, we first focused on having better forecasting performance for ICU patients plus those who died and then on the number of infected persons reported. We gave less importance to the fitting performance, so the model did not overfit to actual data, which have much uncertainty, among other problems discussed in Section 3.1.

From the error criterion given in (12), we define the *best parameter vector*, denoted by $\hat{\boldsymbol{\theta}}_{\bar{k}}$, among the vectors $\hat{\boldsymbol{\theta}}_\ell$ for $\ell = 1, \cdots, M$ that minimizes $E(\hat{\boldsymbol{\theta}}_\ell)$, i.e.,

$$E(\hat{\boldsymbol{\theta}}_{\bar{k}}) := \min_{1 \le \ell \le M} E(\hat{\boldsymbol{\theta}}_\ell) \tag{14}$$

The best parameter vector defined in this way realizes the minimum error of the PBT we implemented and induces the *best model curves*, which privilege the forecasting performance. So, the best model curves are those evaluated at the best parameter vector, $F(t,\hat{\boldsymbol{\theta}}_{\bar{k}})$. In addition, the prediction interval PI produces an envelope of model curves $F(t,\hat{\boldsymbol{\theta}}_{\ell})$ for $\ell = 1,\cdots,M$, quantifying uncertainty around the best model curves $F(t,\hat{\boldsymbol{\theta}}_{\bar{k}})$.

Finally, to assess the impact of vaccination, we compared different immunity durations, the parameter τ_4, from 60 to 240 days, spaced every 30 days. For that, we calculated a weighted mean between the least PBT error $E(\hat{\boldsymbol{\theta}}_{\bar{k}})$ given in (14) with the mean of the **NSE** given in (11b), denoted by $\overline{\mathbf{NSE}}$. We evaluated the best τ_4 as the value minimizing the weighted error, WE, defined by

$$\mathrm{WE} := \frac{2}{3}E(\hat{\boldsymbol{\theta}}_{\bar{k}}) + \frac{1}{3}\overline{\mathbf{NSE}}. \tag{15}$$

We gave more weight (66.67%) to the minimum PBT error and the least weight (only 33.33%) to the mean of the NSE to privilege the FP in (14) over the uncertainty represented by the NSE in (11b), which was within reasonable bounds.

2.2.4. Data processing and actual data limitations

The datasets' sources correspond to the Chilean government's COVID-19 database at the regional level, which we used to track and predict COVID-19 dynamics (10). We used datasets representing different regions of the north, south, and center, including the Metropolitan Region (MR), the major city of which is the country's capital, Santiago de Chile. They correspond to reported infected persons with symptoms confirmed by RT-PCR tests (I_r), recovered cases (R), patients in the ICU, those confirmed dead due to COVID-19 (D), and the size of targeted populations (N). We applied mobile averages with different window sizes to deal with data that were not always reported daily and to smooth out the data.

2.2.4.1. Available actual data limitations

A common problem in processing and modeling epidemiological curve data is the time lag of the information reported, which was not daily but accumulated every 3–5 days in the Chilean case as observed in Figure 1, which depicts various sub-peaks in the official Chilean data curves, specifically in the data of those infected (I), recovered (R), and deceased (D). In addition, the relevant information is only detailed at the national level,

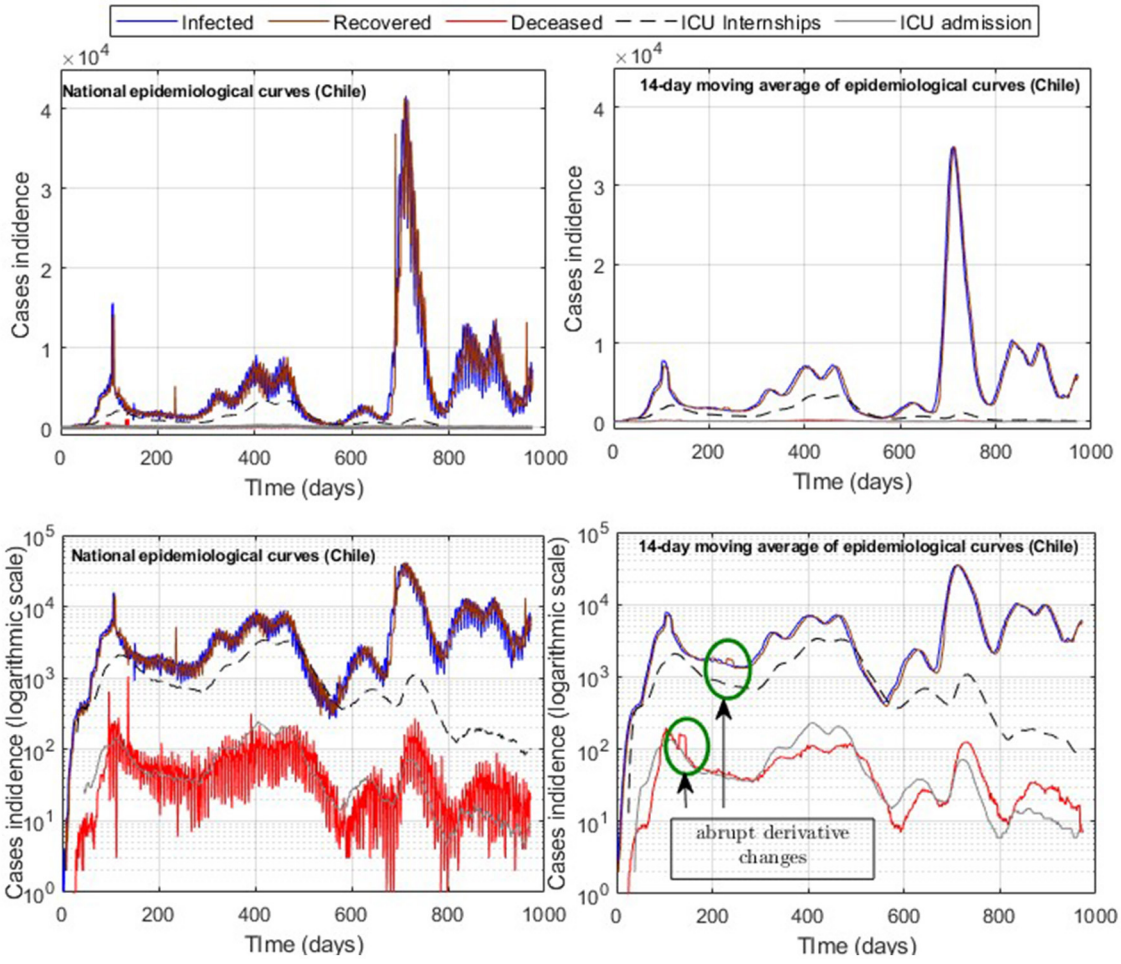


FIGURE 1
Official COVID-19 incidence curves (Chile); raw data and processing with moving averages.

such as daily admissions to the ICU and COVID-19 reinfections according to the vaccination scheme, which was accumulated weekly, and provisional since it was being validated [(10), product 90]. Therefore, using data in constant validation does not ensure a precise replication of the results, unlike I, R, and ICU, which have been maintained without significant variation over time. In addition, in the case of deaths in Chile at the beginning of 2022, the probable cases (without a prior confirmation test for COVID-19) were added to the official data (more than 10,000 cases) for the daily death count. Since the daily distribution of such cases is unknown, in our study, we only considered confirmed death due to COVID-19 reported daily.

Although different epidemiological datasets are available in the official Chilean repository (10), most are designed for statistical studies rather than for modeling studies. The most consistent data (without repetition of cases) is deceased persons, which has a high correlation (Pearson coefficient of 0.8) with hospitalized ICU patients (Figure 2). Therefore, in addition to being relevant data to determine the hospital load, hospitalized ICU patients are of high value in terms of the quality of the available data, from which one can conduct modeling studies at the regional level for the Chilean case. Furthermore, the interest in approximating regional curves and not only the national ones is based on the fact that it allows us to validate our general modeling framework with several datasets and analyze the epidemic in different geographical areas. The previous consideration is relevant to Chile, the longest country in the world, with different climates from north to south, and therefore a good case study for the present work. Finally, we applied our general modeling framework to study the COVID-19 data available (I, R, ICU, and D) and, in the future, we will extend and adapt to any transmissible disease in any country. It is feasible given that in most of the world, it is more viable to track critical cases (those hospitalized in ICUs), deaths, and, to a lesser extent, those who are infected/recovered (the actual total is never reached).

2.2.4.2. Hardware, software, and parallel computing

We used a data science workstation for the careful optimization algorithm implementation with the following features: Intel Core i9 7900x, 10 Cores/20 threads, 128 Gb memory, NVIDIA Titan RTX 24 Gb, and two mobile laptops with Intel Core i7s, 4 Cores/8 threads with 16 Gb memory. We implemented all our calibration codes by using the software Matlab R2022b. In addition, we used the Matlab parallel computing toolbox to speed up the computation with the parpool ("Processes," 20) option on the workstation.

For implementing the PBT (forecasting), we used two mobile laptops equipped with an Intel Core i7 processor with 8 and 12 cores, respectively, with 16 Gb of RAM without parallel computing.

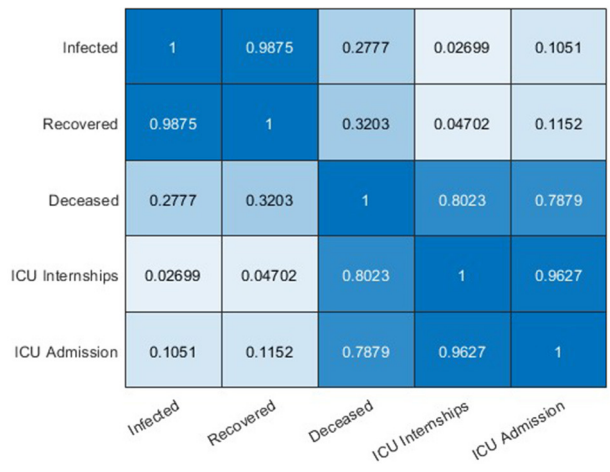


FIGURE 2
Correlations of incidence curves (Chile).

3. Results

3.1. General modeling framework calibration results

3.1.1. Data processing results

Data processing shows that the highest correlated variables are I and R, followed by ICU admission with ICU hospitalized and deceased with ICU hospitalized (Figure 3). The two latter are the most relevant indicators of the impact of the pandemic on the health system, which justifies our choice of privileging the forecasting performance in U, as made in Equation (12).

In addition, the optimization results strongly depend on the parameters' initial values and the fitted data quality. By applying moving averages of 14 days, we reduced abrupt slope changes of the actual epidemiological curves (Figure 1), improving results with information loss of less than 2% of incidences for both infected and ICU. In addition, using moving windows combined with the cumulative distribution curves (Figure 3) reduces the number of steps of the Matlab dde23 solver (for non-linear DDE systems) because the slopes are always positive and not as steep as for daily curves. In this way, our method allows for calibrating parameters with accumulated and daily curves, where the cumulative data is helpful in efficiently fitting from the pandemic's beginning to any subsequent location and day to perform block tracking.

3.1.2. Calibration results

The percentage decrease technique, described in Section 2.2.2.4, ensures the overall method stability and convergence by preventing the NaN appearance (NaN means "Not a Number") when overflow occurs (computational numerical limit exceeded). NaNs are propagated in the model's history function, making the optimization solver lsqnonlin require more time to find new feasible points and often diverge. Using different ω values in $\boldsymbol{\theta}^{(0)} := \omega \cdot \mathrm{UB}_{\boldsymbol{\theta}}$, one may generate suitable initial curves for I and U below the epidemiological curves. For small ω values ($\omega \in [1e{-}4, 1e{-}2]$), the initial curve will approximately be a straight line, obtained with government data, and greater than zero. By applying this simple but effective technique, our careful optimization algorithm implementation ensures stability and convergence for the dde23 solver and the lsqnonlin optimization subroutine within the range of suitable parameters, as described in Section 2.2.2.3.

Furthermore, to speed up the computation, we performed a one-step initial optimization defining the real positivity rate $f(t) = 1$. This was the same as taking $I(t) = I_r(t)$, i.e.,

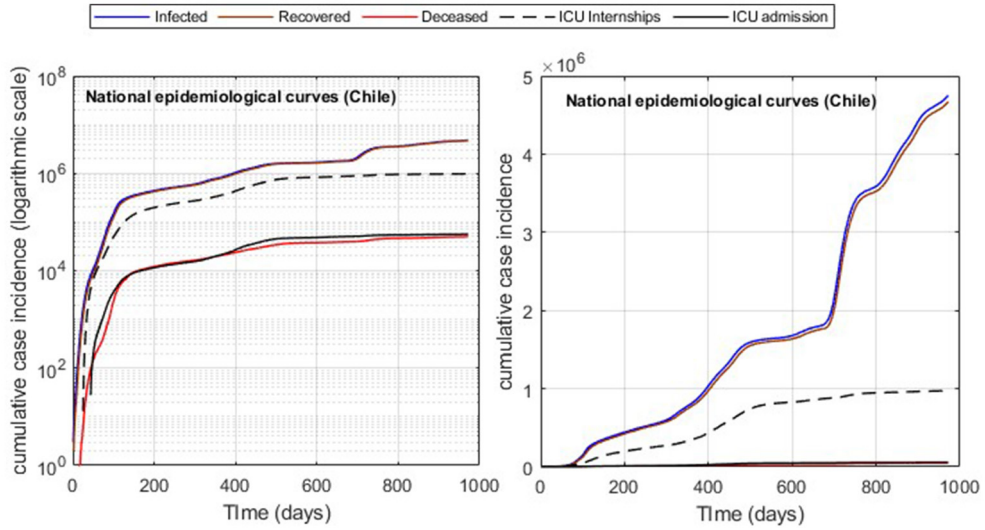


FIGURE 3
Cumulative incidence curves (Chile). Logarithmic and raw data, respectively.

Table 2 Model results for different ω.

Id	Fraction ω	Curve/window	$\overline{\text{RE}_I}$	$\overline{\text{RE}_R}$	$\overline{\text{RE}_{ICU}}$	$n_\beta + 1$	t (days)	UB_{τ_4}
MR.1, Figure 4	1e-03	cum/2 weeks	$3.2808e-01$	$4.1021e-01$	$1.8342e-01$	20	[30, 950]	240
MR.2, Figure 5	1e-03	cum/2 weeks	$8.2663e-02$	$1.4144e-01$	$1.99823-02$	20	[30, 250]	240
MR.3, Figure 6	1e-03	cum/2 weeks	$4.5210e-02$	$1.6722e-01$	$9.4964e-03$	20	[30, 250]	50
VAL.1, Figure 7	1e-03	cum/3 weeks	$1.0394e+00$	$6.9195e-01$	$7.9057e-01$	20	[250, 950]	240
ANT.1, Figures 8C, D	3.5e-03	daily/3 weeks	$1.8747e-01$	$1.0747e+00$	$8.2074e-02$	20	[250, 600]	240
MAG.1, Figure 9	2.5e-04	cum/2 weeks	$9.9075e-01$	$3.2812e+00$	$3.4139e+00$	20	[100, 950]	240

all the actual infected are reported on the entire curve and storing the resulting ω as a checkpoint. Then, when required to optimize the curve over any time interval, the obtained ω was used as the initial value, starting the computation closer to the optimum.

In Table 2, we present the optimization results for several datasets located throughout all of Chile: the Metropolitan region, Valparaíso region (center), Antofagasta region (northern), and Magallanes region (southern), designated by the identifiers MR, VAL, ANT, and MAG in the Id column. We plotted the corresponding curves in Figures 4–9. In addition, Table 2 shows the results concerning a study on the immunity duration in different time intervals to assess vaccination. We obtained the calibration results by varying the upper bound UB_{τ_4} of the parameter τ_4 for a dataset that contains the MR's first wave; see rows MR.2-MR.3 in Table 2 and Figures 5, 6. The interpretation was that the least mean relative error for I and U, $\overline{\text{RE}_I}$ and $\overline{\text{RE}_U}$, implied that the corresponding τ_4 value was the most probable for the respective dataset. This τ_4 variation is helpful for the analysis, simulation, and evaluation of epidemiological scenarios where, for example, a better fit for τ_4 larger means a high vaccine immunity duration.

From Table 2, for example, by decreasing the upper bound of τ_4, comparing the results between MR.2 and MR.3 (through $\overline{\text{RE}_I}$ and $\overline{\text{RE}_U}$), we observe that τ_4 is small, which could be interpreted as correct, since in the MR's first wave there was no vaccination and therefore no immunity due to it.

General epidemiological modeling with cumulative data allows us to fit the complete epidemiological curve for analytical purposes (MR.1 in Table 2) with slightly less precision. In some cases, it was necessary to increase the moving window size to handle abrupt changes of derivative in the curves or optimize with other fractions of the upper bounds (different values of ω). Although the error obtained may increase, these approximations (Val.1 in Table 2) should be considered since they may indicate a data problem or a change in the epidemiological scenario. The fact that these cases of higher error occur only for the data from those infected is another reason to focus more on the number of ICU patients plus deceased.

An example of the cases explained above is the Antofagasta region, the results from which are quantified with the ANT.1 identifier in the Id column in Table 2 and depicted in Figure 8A. We observed a wave not reported in the data for $t \in [250, 950]$. In addition, Figure 8C depicts that the model fits data for $t \in [250, 600]$, with low error for I and U, even if the actual infected (calculated by the model) was much higher than reported.

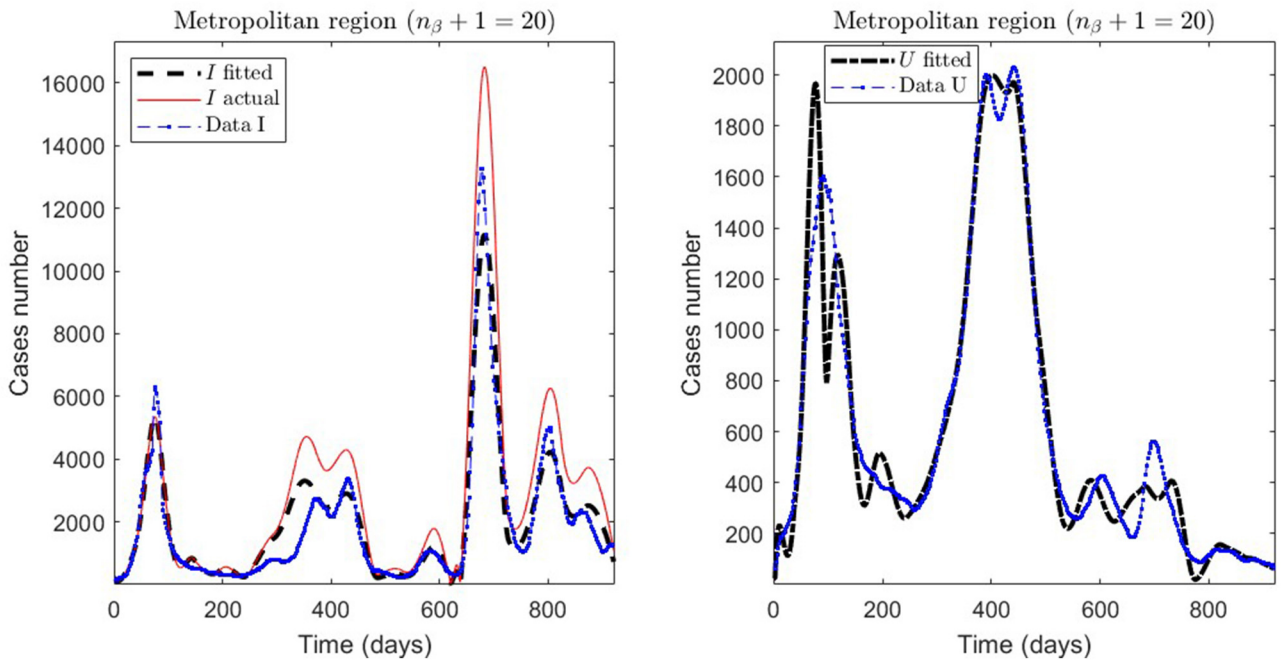


FIGURE 4
Fitted incidence curves in the Metropolitan region, Chile, 1 April 2020 to 8 October 2022.

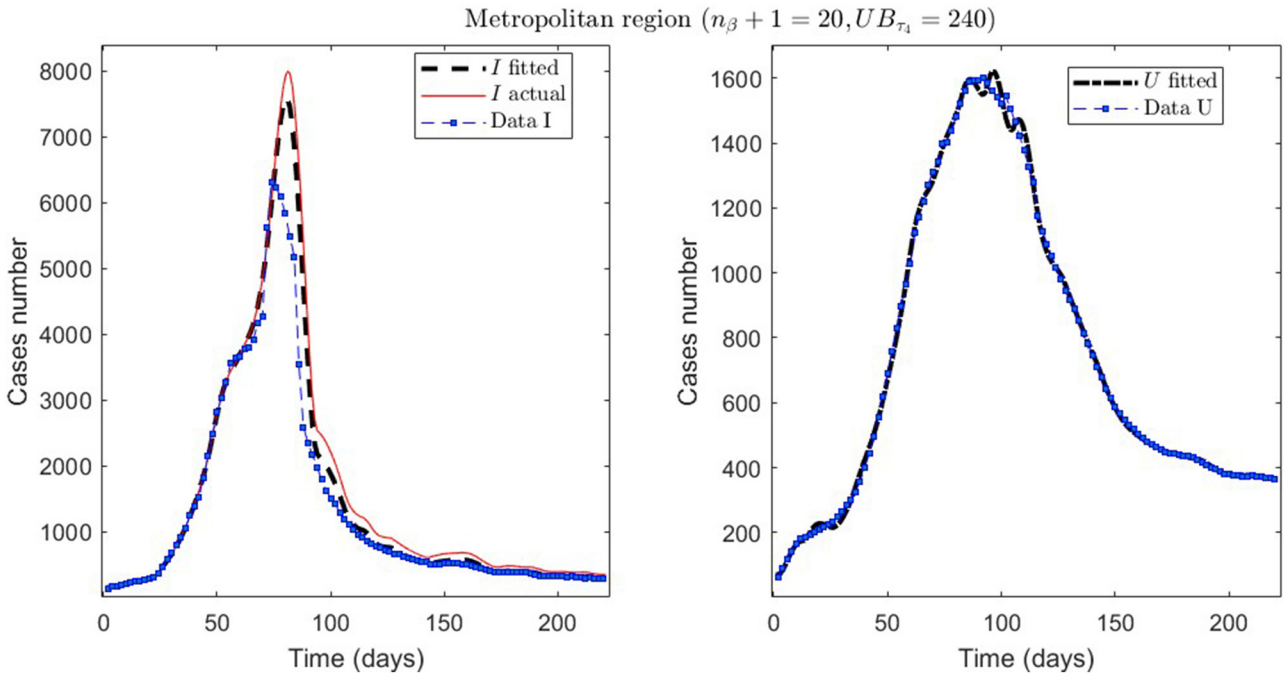


FIGURE 5
The first epidemiological wave in the Metropolitan region, Chile. 1 April 2020 to 7 November 2020; $UB_{\tau_4} = 240$ (unrestricted immunity duration).

Finally, we present the case of the Magallanes region (MAG.1 in Table 2), a zone located in the extreme south of Chile with very few inhabitants and, therefore, little data. Despite this, the model also fits curve U reasonably, despite a more significant error, where the relevance is to interpret the curve's trend.

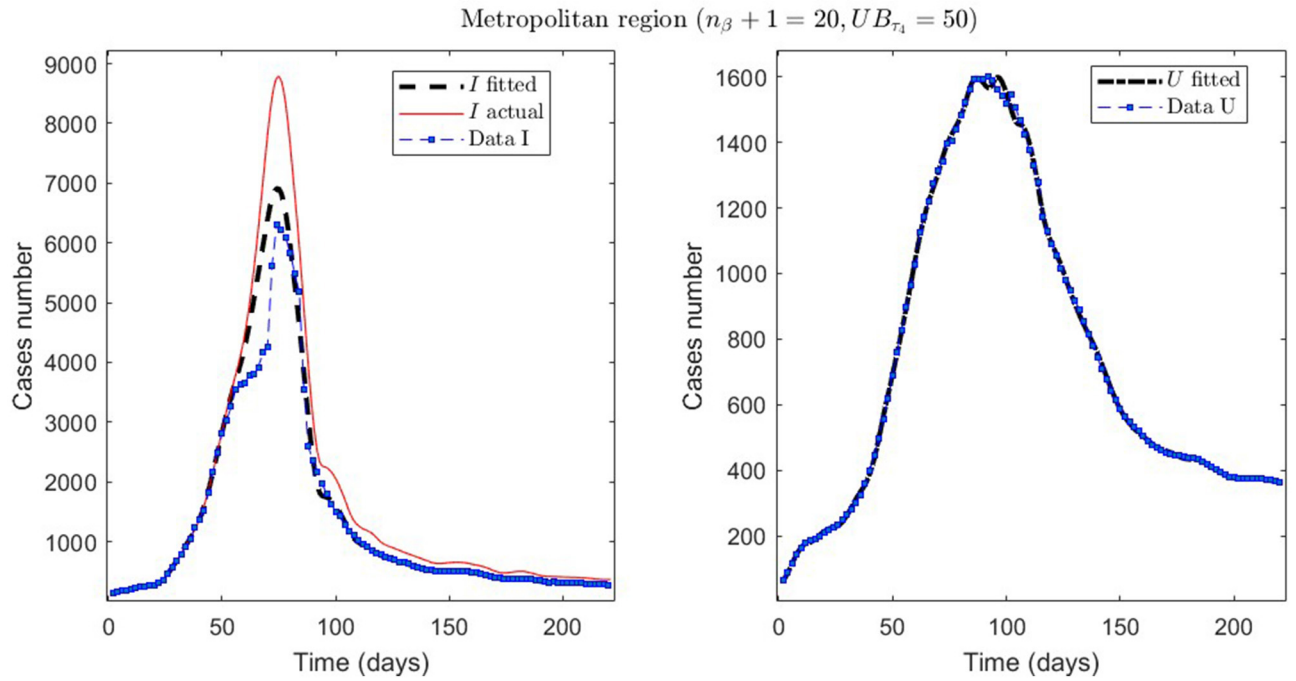


FIGURE 6
The first epidemiological wave in the Metropolitan region, Chile. 1 April 2020 to 7 November 2020; $UB_{\tau_4} = 50$ (immunity duration restriction).

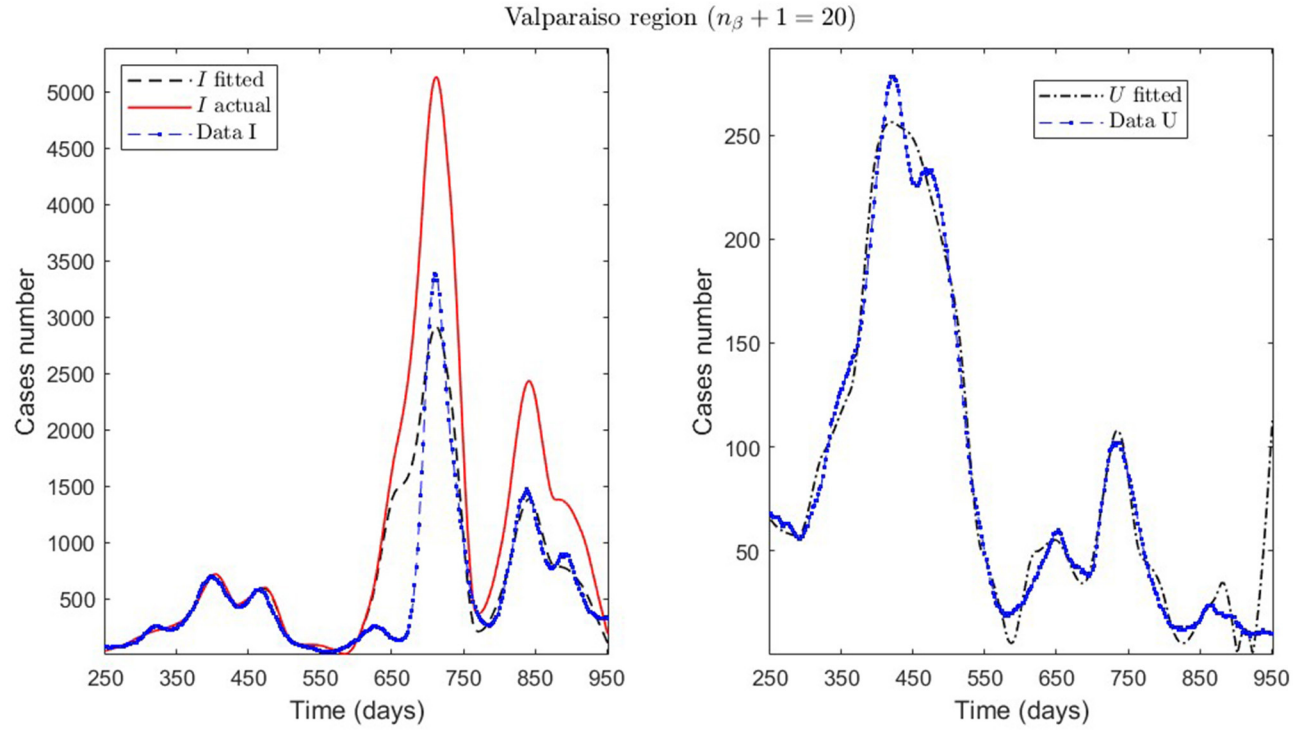


FIGURE 7
Fitted incidence curves in the Valparaíso region, Chile. 7 October 2020 to 8 October 2022; $UB_{\tau_4} = 240$ (unrestricted immunity duration).

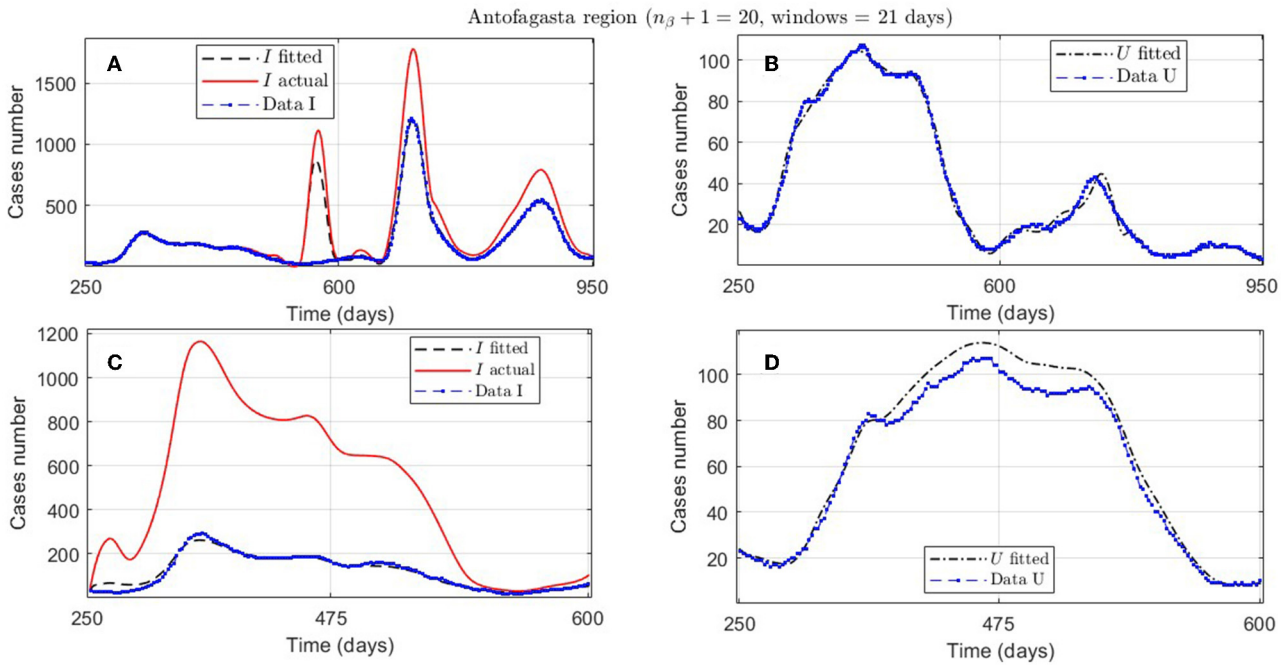


FIGURE 8
Fitted incidence curves in the Antofagasta region, Chile; $UB_{\tau_4} = 240$ (unrestricted immunity duration). **(A, B)** 7 November 2020 to 8 October 2022; **(C, D)** 7 November 2020 to 23 October 2021.

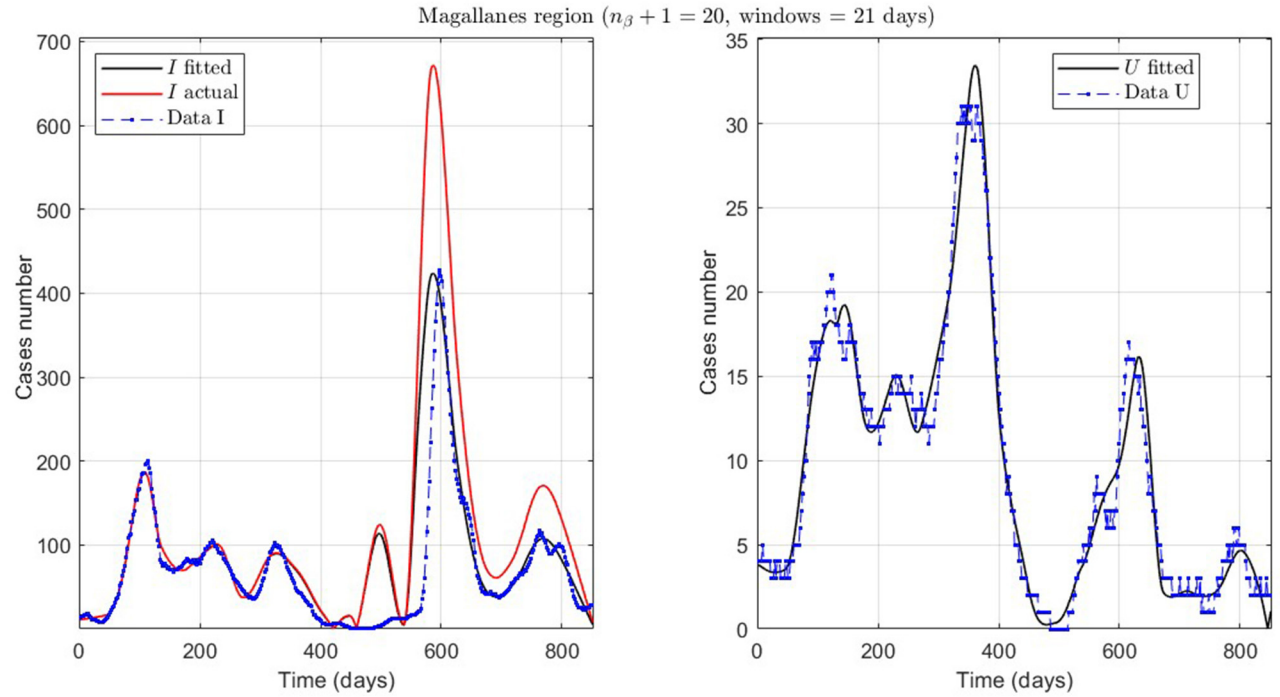


FIGURE 9
Fitted incidence curves in the Magallanes region, Chile. 10 June 2020 to 8 October 2022; $UB_{\tau_4} = 240$ (unrestricted immunity duration).

3.2. Parametric bootstrap technique results

In this section, we want to show the predictive capabilities of our general modeling framework. For that, we applied the PBT for the 280–530 time interval in the Metropolitan region (MR), which encompasses the MR's second wave, and the prediction period was 12 weeks.

First, we demonstrate the results obtained using the careful parameter optimization for fitting general epidemiological modeling to the actual data by minimizing the sum of squares (7) (step 1 of the algorithm described in Section 2.2.3). A code run with this method with 100 iterations took around 14 min in an Intel Core i7 processor with eight cores and 16 Gb of RAM.

We summarize the numerical results in Table 3 and depict the least-squares fitted model curves $F(t_j, \hat{\boldsymbol{\theta}})$ (Equation 9) in Figure 10. The corresponding error evaluated in the non-linear LSE $\hat{\boldsymbol{\theta}}$, according to Equation (12), is $E(\hat{\boldsymbol{\theta}}) = 4.3387e-01$.

Second, we show the model results obtained through the PBT to achieve a better fitting and forecasting performance (the entire algorithm is described in Section 2.2.3). The PBT is computationally intensive since the code runs took around 24/16 h on one/two high-performance laptops, as described in Section 2.2.4.2. Using the two laptops, we calculated 50 parameter estimates to obtain the $M = 100$ bootstrap parameter realizations for fitting general epidemiological modeling to every synthetic dataset generated by a normal distribution with moderate variance relative to the least-squares curves $F(t_j, \hat{\boldsymbol{\theta}})$ [Equation (9)], as explained in Section 2.2.3. Then, we constructed the best parameter estimate [$\hat{\boldsymbol{\theta}}_{\bar{k}}$ defined in (14)] and the PI [$\text{LB}_{\hat{\theta}}$ and $\text{UB}_{\hat{\theta}}$ defined in (11)]. With these estimates, we plotted the respective curves for infected persons I_r and the sum of ICU patients and those who were reported dead U_r.

Figure 11 depicts the model results for synthetic datasets constructed as explained above. The minimum PBT error was $E(\hat{\boldsymbol{\theta}}_{\bar{k}}) = 2.8775e-01$, calculated according to Equation (14). Finally, the parameter estimates uncertainty, corresponding to the curves plotted in Figure 11, ranged between 0.02 and 18.92% with a mean of 4.61% measured in normalized standard error (NSE) in percent, as defined in Equation (11b).

Table 3 Least-squares model results.

Criterion \ Variable	I	U
FP	6.3983e-01	4.3173e-01
RMSE	4.3539e-01	3.3227e-02

3.3. Assessing the impact of vaccination

We compared different values of the parameter τ_4 in the range of 60–240 days to assess the impact of vaccination on immunity. For τ_4 values spaced every 30 days, and ranging between 60 and 240 days, we obtained the results summarized in Table 4, which shows the minimum PBT error $E(\hat{\boldsymbol{\theta}}_{\bar{k}})$ given in (14), the mean of the NSE ($\overline{\text{NSE}}$) given in (11b), and the weighted error WE given in (15).

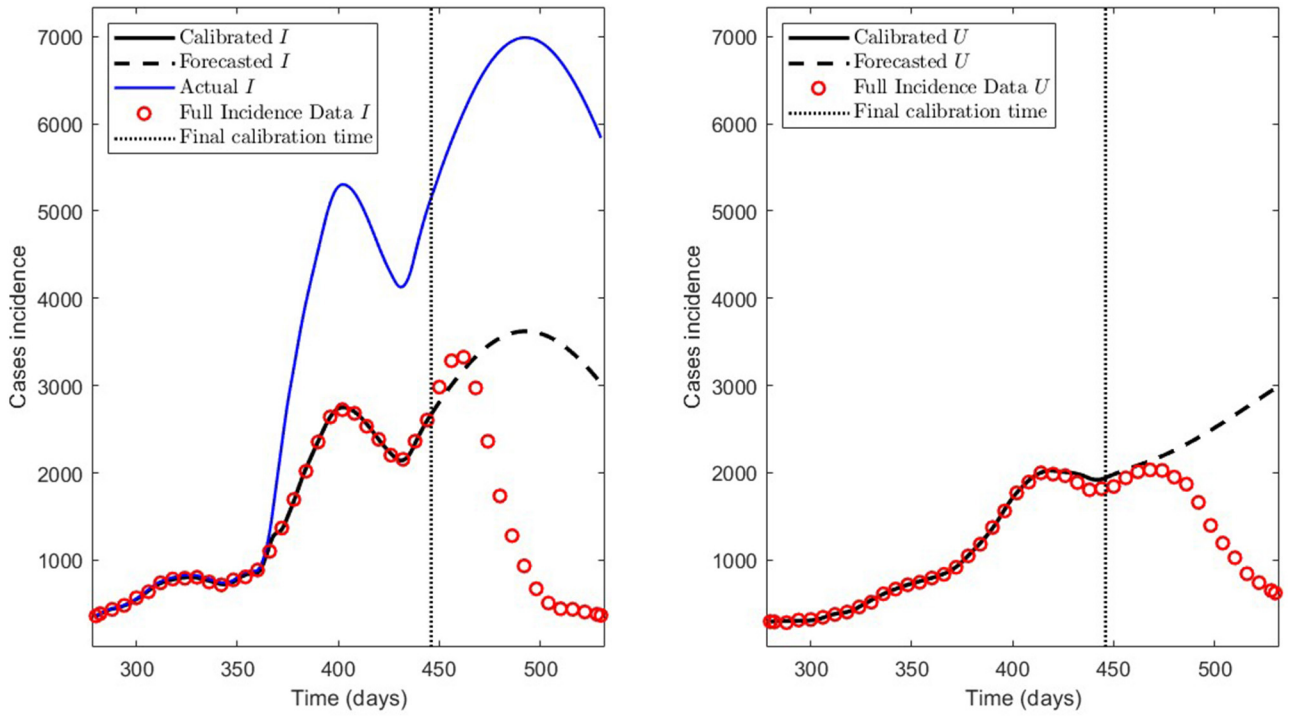


FIGURE 10
Incidence curves obtained by least-squares. 7 December 2020 to 14 August 2021, encompassing the Metropolitan region's second wave in Chile.

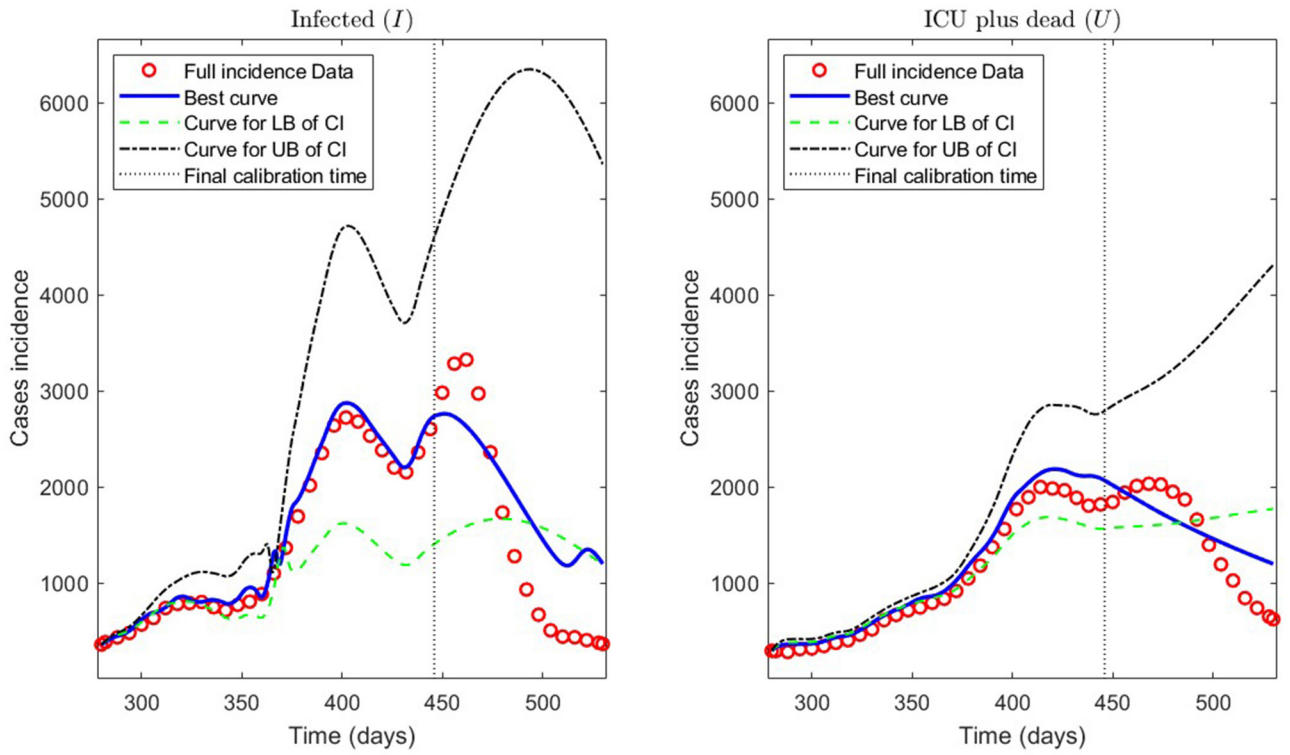


FIGURE 11
Incidence curves obtained by the PBT. 7 December 2020 to 14 August 2021, encompassing the Metropolitan region's second wave in Chile.

From Table 4, we observe that the best value for τ_4 was $\tau_4 = 60$ days since it yielded the least weighted error WE.

Table 4 Model results for different values of τ_4.

τ_4	$E\left(\hat{\theta}_{\bar{k}}\right)$	$\overline{NSE}$	WE
60	**2.7309e-01**	**4.9037e-02**	**1.9840e-01**
90	2.8199e-01	4.9866e-02	2.0462e-01
120	2.9898e-01	5.3788e-02	2.1725e-01
150	2.9882e-01	6.2228e-02	2.1995e-01
180	2.8646e-01	5.4880e-02	2.0926e-01
210	3.4228e-01	5.1622e-02	2.4539e-01
240	3.5367e-01	6.2833e-02	2.5672e-01

τ_4 is the mean immunity duration by vaccination; see Table 1.
$E(\hat{\theta}_{\bar{k}})$ is the minimum error of the parametric bootstrap technique (PBT); see Equation (14).
$\overline{\mathbf{NSE}}$ is the mean normalized standard error (NSE); see Equation (11b), and
WE is the weighted error; see Equation (15).

4. Discussion

Concerning quantitative tracking of COVID-19 dynamics, we can calibrate any dataset with our general modeling framework. In effect, we show the framework's calibration capabilities through several examples for different regions of Chile; see Table 2 and the corresponding plotted curves in Figures 4–9. In addition, we corroborated that immunity duration was short ($\tau_4 \leq 50$ days, as shown in row MR.3 in Table 2) during the Metropolitan region's first wave (when there was no vaccination yet). Consequently, our general modeling framework provides a flexible tool for studying the dynamics of any transmissible disease and assessing vaccination, despite the vaccination deficiencies and data limitations discussed in Sections 2.1.2.1, 2.2.4.1. To do so, it suffices to have time-series observations for the number of infected persons and ICU patients and to find a suitable range of initial parameters meaningful from the epidemiology viewpoint. Then, the percentage decrease technique allows us to find a unique range of parameters proper for every studied dataset, which provides overall method numerical stability and convergence, as shown in this article.

Concerning the framework's predictive capabilities, we applied the PBT to a dataset encompassing the Metropolitan region's second wave. The results shown in Table 3 and depicted in Figure 10, show that fitting the model to processed data by least-squares produces a forecasting performance for variables I and U that could be better (even more for U), despite their respective fitting performance (RMSE) being excellent. From the PBT implementation, we found that the parameter uncertainty range, evaluated through the percentage NSE, needed to have reliable parameter estimates (between 0.02 and 18.92% with a mean of 4.61%), so the number of bootstrap samples $M = 100$ was suitable. In addition, Figure 11 shows an outstanding forecasting performance for variables I and U (even more for U) with an error of 1.5 times less than for the usual least-squares method ($E(\hat{\boldsymbol{\theta}}) \approx 1.5E(\hat{\boldsymbol{\theta}}_{\bar{k}})$). Therefore, our general epidemiological

framework can accurately predict the ICU curve trend in the medium term (12 weeks). However, a limitation to achieving such a good prediction performance is that the PBT is costly from a computational viewpoint.

In Figure 11, we observe that the segmented curves (green lines) are more significant than the filled curves (blue lines) in specific time intervals, despite that the parameter vector of the firsts [evaluated at the lower bound of the PI; Equation (11a)] is smaller than the one of the seconds [evaluated at the best parameter vector; Equation (14)]. It may happen since, according to Equations (5b) and (5d), the curves for variables I and U will be significant if the $\gamma_{IR}(t)$, $\gamma_{IU}(t)$, and $\gamma_{UR}(t)$ rates are close to zero in some time intervals (which is the case of the segmented curves).

Concerning vaccination assessment relative to its immunity duration, from Table 4, we infer that $\tau_4 > 60$ values do not adapt to the dataset encompassing the MR's second wave. Therefore, the immunity duration should be less than 60 days. This exciting result implies a short immunity duration during the MR's second wave, which is not surprising since vaccination at that time was not yet widespread. In addition, this result would corroborate that the vaccination effect is not as significant as the immunity provided by natural infection, as discussed in Section 2.1.2.1. Indeed, the third wave magnitude (the most prolonged and steepest so far) shows that vaccination was relatively ineffective regarding protection against infection before it. However, one may think that vaccination manifests a positive effect by the time of the third wave, which we can infer from the fact that the ICU patient and death data are low compared to the infected data during that wave.

Our results rely on the piecewise linear reconstruction of time-varying parameters γ_{IR}, γ_{SR}, γ_{RS}, γ_{IU}, β, and γ_{UR}. We could improve this arbitrary choice by assuming the mechanical laws of the transmission rate (β) or other rates as, for example, in (21).

5. Conclusions

This work attempts to implement a holistic and general modeling framework for quantitative tracking of the dynamics of any transmissible disease, focusing on accurately predicting the ICU curve trend in the medium term and assessing vaccination.

Implementing a careful optimization algorithm, we obtained outstanding results concerning quantitative tracking of COVID-19 dynamics for several processed datasets representing different regions of Chile and assessing vaccination. In addition, a parametric bootstrap technique allowed us to predict the ICU curve trend in the medium term accurately and assess vaccination. As a result, the scientific community could adapt our general modeling framework to evaluate the impact of combined vaccination and NPI strategies for COVID-19 or any transmissible disease in any country and help visualize the potential effects of implemented plans by policymakers.

In conclusion, the two main lessons are that we must be prepared to face another pandemic like COVID-19 and that it is more important to make more effort to follow up the ICU patients, which are highly correlated with dead confirmed by COVID-19.

To tackle the first lesson, in future work, we want to improve the computational cost of the parametric bootstrap technique or use another technique more efficiently. The aim would be to reconstruct epidemiological curves to predict the combined NPIs and vaccination policies' impact on the ICU curve trend in real time, providing scientific evidence to help anticipate policymakers' decisions.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/MinCiencia/Datos-COVID19.

Author contributions

PC contributed to conceptualization, formal analysis, investigation, methodology, project administration, software, supervision, validation, writing—original draft, review, and editing. OR-D contributed to conceptualization, formal analysis, investigation, methodology, software, validation, and writing—review and editing. CC contributed to theoretical analysis, supervision, and writing—review and editing. All authors contributed to the article and approved the submitted version.

Funding

The Centre for Biotechnology and Bioengineering (CeBiB), Grant Number ANID PFBasal-01, partially supported PC and CC. In addition, PC was partially supported by the regular research project DIUBB 2120432 IF/R. Finally, CC was partially supported by the ACE210010 and CMM PIA AFB170001 projects. The funders had no role in study design, data collection, analysis, publication decision, or manuscript preparation.

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.

Publisher's note

References

1. Al-Tuwairqi SM, Al-Harbi SK. A time-delayed model for the spread of COVID-19 with vaccination. *Sci Rep.* (2022) 12:19435. doi: 10.1038/s41598-022-23822-5

2. Zhenzhen L, Yongguang Y, YangQuan C, Guojian R, Conghui X, Shuhui W. Stability analysis of a nonlocal SIHRDP epidemic model with memory effects. *Nonlinear Dyn.* (2022) 109:121–41. doi: 10.1007/s11071-022-07286-w

3. Ghosh S, Volpert V, Banerjee M. An epidemic model with time delay determined by the disease duration. *Mathematics.* (2022) 10:2561. doi: 10.3390/math10152561

4. Zhai S, Luo G, Huang T, Wang X, Tao J, Zhou P. Vaccination control of an epidemic model with time delay and its application to COVID-19. *Nonlinear Dyn.* (2021) 106:1279–92. doi: 10.1007/s11071-021-06533-w

5. Cumsille P, Oscar Rojas-Diaz, de Espanas PM, Verdugo-Hernandez P. Forecasting COVID-19 Chile' second outbreak by a generalized SIR model with constant time delays and a fitted positivity rate. *Math Comput Simul.* (2022) 193:1–18. doi: 10.1016/j.matcom.2021.09.016

6. Cumsille P, Coronel A, Conca C, Quiñinao C, Escudero C. Proposal of a hybrid approach for tumor progression and tumor-induced angiogenesis. *Theoret Biol Med Modell.* (2015) 12:13. doi: 10.1186/s12976-015-0009-y

7. Badillo G, Cumsille P, Segura-Ponce L, Pataro G, Ferrari G. An efficient optimization methodology of respiration rate parameters coupled with transport properties in mass balances to describe modified atmosphere packaging systems. *Inverse Prob Sci Eng.* (2020) 28:1361–83. doi: 10.1080/17415977.2020.1717488

8. Cumsille P, Godoy M, Gerdtzen ZP, Conca C. Parameter estimation and mathematical modeling for the quantitative description of therapy failure due to drug resistance in gastrointestinal stromal tumor metastasis to the liver. *PLoS ONE.* (2019) 14:e0217332. doi: 10.1371/journal.pone.0217332

9. Chowell G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: a primer for parameter uncertainty, identifiability, and forecasts. *Infect Dis Modell.* (2017) 2:379–98. doi: 10.1016/j.idm.2017.08.001

10. Cornejo S, Arancibia D, Molina F, Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, Frías IA. *Official Chilean Repository of COVID19 Data.* (2022). Available online at: https://github.com/MinCiencia/Datos-COVID19

11. Miao H, Xia X, Perelson AS, Wu H. On Identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev.* (2011) 53:3–39. doi: 10.1137/090757009

12. Müller S, Lu J, Kügler P, Engl H. *Parameter Identification in Systems Biology: Solving ill-Posed Inverse Problems Using Regularization.* Johann Radon Institute for Computational and Applied Mathematics (RICAM); Austrian Academy of Sciencies (ÖAW) (2008). Available online at: https://www.ricam.oeaw.ac.at/files/reports/08/rep08-25.pdf

13. Moore S, Hill EM, Tildesley MJ, Dyson L, Keeling MJ. Vaccination and non-pharmaceutical interventions for COVID-19: a mathematical modelling study. *Lancet Infect Dis.* (2021) 21:793–802. doi: 10.1016/S1473-3099(21)00143-2

14. Pérez-Alós L, Armenteros JJA, Madsen JR, Hansen CB, Jarlhelt I, Hamm SR, et al. Modeling of waning immunity after SARS-CoV-2 vaccination and influencing factors. *Nat Commun.* (2022) 13:1614. doi: 10.1038/s41467-022-29225-4

15. GSM-covid19CL. *Github Repository Code Paper* (2022). Available online at: https://github.com/neurovisionhub/GSM-covid19CL

16. Banks HT, Davidian M, Samuels JR, Sutton KL. An inverse problem statistical methodology summary. In: Chowell G, Hyman JM, Bettencourt LMA, Castillo-Chavez C, editors. *Mathematical and Statistical Estimation Approaches in Epidemiology.* Dordrecht: Springer (2009). p. 249–302. doi: 10.1007/978-90-481-2313-1_11

17. Shampine LF, Thompson S. Solving DDEs in Matlab. *Appl Num Math.* (2001) 37:441–58. doi: 10.1016/s0168-9274(00)00055-6

18. WHO. *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)* (2019). Available online at: https://www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19)

19. Dan JM, Mateus J, Kato Y, Hastie KM, Yu ED, Faliti CE, et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science.* (2021) 371:eabf4063. doi: 10.1126/science.abf4063

20. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap.* 1st ed. New York, NY: Chapman and Hall/CRC (1994). doi: 10.1201/9780429246593

21. Córdova-Lepe F, Vogt-Geisse K. Adding a reaction-restoration type transmission rate dynamic-law to the basic SEIR COVID-19 model. *PLoS ONE.* (2022) 17:e0269843. doi: 10.1371/journal.pone.0269843

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 11 April 2023
DOI 10.3389/fpubh.2023.1085991

OPEN ACCESS

EDITED BY
Theodore Gyle Lewis,
Naval Postgraduate School, United States

REVIEWED BY
Orvalho Augusto,
University of Washington, United States
Simon Grima,
University of Malta, Malta

*CORRESPONDENCE
Bruno Enagnon Lokonon
✉ brunolokonon@gmail.com

†These authors have contributed equally to this work and share last authorship

SPECIALTY SECTION
This article was submitted to Infectious Diseases: Epidemiology and Prevention, a section of the journal Frontiers in Public Health

RECEIVED 31 October 2022
ACCEPTED 27 March 2023
PUBLISHED 11 April 2023

CITATION
Lokonon BE, Montcho Y, Klingler P, Tovissodé CF, Glèlè Kakaï R and Wolkewitz M (2023) Lag-time effects of vaccination on SARS-CoV-2 dynamics in German hospitals and intensive-care units.
Front. Public Health 11:1085991.
doi: 10.3389/fpubh.2023.1085991

Lag-time effects of vaccination on SARS-CoV-2 dynamics in German hospitals and intensive-care units

Bruno Enagnon Lokonon[1,2]*, Yvette Montcho[1,2], Paul Klingler[2], Chénangnon Frédéric Tovissodé[1], Romain Glèlè Kakaï[1†] and Martin Wolkewitz[2†]

[1]Laboratoire de Biomathématiques et d'Estimations Forestières, Université d'Abomey-Calavi, Cotonou, Benin, [2]Faculty of Medicine and Medical Center, Institute of Medical Biometry and Statistics, University of Freiburg, Freiburg im Breisgau, Germany

Background: The Efficacy and effectiveness of vaccination against SARS-CoV-2 have clearly been shown by randomized trials and observational studies. Despite these successes on the individual level, vaccination of the population is essential to relieving hospitals and intensive care units. In this context, understanding the effects of vaccination and its lag-time on the population-level dynamics becomes necessary to adapt the vaccination campaigns and prepare for future pandemics.

Methods: This work applied a quasi-Poisson regression with a distributed lag linear model on German data from a scientific data platform to quantify the effects of vaccination and its lag times on the number of hospital and intensive care patients, adjusting for the influences of non-pharmaceutical interventions and their time trends. We separately evaluated the effects of the first, second and third doses administered in Germany.

Results: The results revealed a decrease in the number of hospital and intensive care patients for high vaccine coverage. The vaccination provides a significant protective effect when at least approximately 40% of people are vaccinated, whatever the dose considered. We also found a time-delayed effect of the vaccination. Indeed, the effect on the number of hospital patients is immediate for the first and second doses while for the third dose about 15 days are necessary to have a strong protective effect. Concerning the effect on the number of intensive care patients, a significant protective response was obtained after a lag time of about 15–20 days for the three doses. However, complex time trends, e.g. due to new variants, which are independent of vaccination make the detection of these findings challenging.

Conclusion: Our results provide additional information about the protective effects of vaccines against SARS-CoV-2; they are in line with previous findings and complement the individual-level evidence of clinical trials. Findings from this work could help public health authorities efficiently direct their actions against SARS-CoV-2 and be well-prepared for future pandemics.

KEYWORDS

delayed effects, vaccination, non-pharmaceutical interventions (NPIs), linear lag models, COVID-19, policy decisions

1. Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that emerged in China in late 2019 has caused major public health concerns and continues to spread worldwide (1–3). There have been a total of 614 million confirmed cases globally, with over 6 million deaths reported, as of August 31, 2022, WHO (4). In Germany, the first COVID-19 case was reported on January 27, 2020, in Bavaria and by March 1, 2020, more than 100 cases were reported (5). Non-Pharmaceutical Interventions (NPIs) have quickly been promoted by the federal government including schools, kindergartens, universities, borders for travelers closing as well as national curfew and contact ban (6). The infection rate decreased following these measures, however, in mid-July 2020, the number of cases started to rise again due to relaxation (7). NPIs have been sufficiently effective in curtailing and mitigating the burden of the pandemic during at least its first waves (6, 7), however, some of them such as containment and travel ban could not be maintained for long times. It was then believed that the use of vaccines combined with some control measures may be necessary to effectively curtail and eliminate COVID-19 (8).

In Germany, the vaccination program began on December 27, 2020, and as of August 31, 2022, 77.66% of all German population have been fully vaccinated (9). The most used vaccines in Germany are BioNTech (95% of efficacy), Moderna (94.1%), AstraZeneca (67%), and Johnson & Johnson (67%) (10–12).

The effect of vaccines is manifold as they act on the individual as well as the population level (13). On the individual level, vaccines aim to reduce the risk of acquiring the infection and transmission but also the clinical consequences once infected. The gold standard study designs to assess vaccine efficacy are randomized placebo-controlled trials. In addition, cohort and case-control studies are used to measure the vaccine effectiveness during real-world conditions (13). For the SARS-CoV-2 pandemic, several studies have shown vaccines efficacy and effectiveness (14–17).

Contrary to cohort and case-control studies, ecological or trend studies compare results on population level over time with varying vaccine coverage (13). The classical approach in ecological studies is to extrapolate from time trends before vaccine introduction, thus creating counterfactual settings which are essential for causal inference. However, the variants of SARS-CoV-2 and their different impact on the pandemic dynamic have made extrapolation extremely difficult.

Moreover, several mathematical models have been developed to predict and assess the impact of vaccination on the transmission dynamics of COVID-19. Gnanvi et al. (18) performed a systematic and critical review on the reliability of predictions of the modeling techniques on COVID-19 dynamics. Dashtbali and Mirzaie (19) used a Susceptible, Exposed, Infected, Hospitalized, Recovered, and Death compartmental model and found that, in the German population, the number of infected cases at the epidemic peaks decreases by increasing the vaccine coverage. Wollschläger et al. (20) applied a multivariable logistic regression on data from the German federal state of Rhineland-Palatinate and concluded that vaccination coverage was associated both with a reduction in the age-groups proportion of COVID-19 fatalities and of reported SARS-CoV-2 infections. Braun et al. (21) developed an effect model based on the Batman-SIZ algorithm for modeling the effect of vaccination on the course of the pandemic in Germany. They obtained that, the effect of vaccination in reducing the daily number of new infections, the total number of infections and the occupancy of intensive-care facilities in hospitals is proportional to the speed with which the target population are vaccinated. Springer et al. (22) used linear regression on the 4th corona wave in Germany and showed that there is a negative correlation between the vaccination rate and the infection incidence. Campos et al. (23) used a membrane computing model for simulating the efficacy of vaccines on the epidemiological dynamics of SARS-CoV-2. They obtained that generalized vaccination of the entire population (all ages) added little benefit to overall mortality rates. However, elderly-only vaccination, even without general interventions directed to reduce population transmission, is sufficient for dramatically reducing mortality. Stiegelmeier et al. (24) proposed a p-fuzzy system in order to model the COVID-19 epidemic evolution under the effect of vaccination in Brazil. They concluded that the level of infestation tends to decrease as the number of people vaccinated increases. Sepulveda et al. (25) constructed a mathematical model based on a nonlinear system of delayed differential equations to investigate the qualitative behavior of the COVID-19 pandemic under an initial vaccination program. They found that if the basic reproduction number is less than one and the time delays are less than some critical threshold, then the disease-free equilibrium is locally stable. Thus, if public health authorities are able to reduce transmission rates and increase vaccination rates, the burden of the COVID-19 pandemic can be reduced.

Despite the contributions of these studies, they showed some limitations. First, the classic mathematical models of epidemiological prediction are quite useful, but deterministic, demonstrating only the average behavior of the epidemic, which makes it difficult to quantify uncertainty (26). Second, the effect of vaccination on COVID-19 reported data may not be linear. Third, vaccination may also show effects that are delayed in time, requiring assessment of the temporal dimension of the exposure-response relationship (27). In addition, the previous studies ignore the seasonal patterns of COVID-19 and the long-term trends in the data. Indeed, the main challenge of modeling the effects of exposure like vaccination on COVID-19 reported data lies in the additional temporal dimension needed to express this relation, as the effects depend on both intensity and timing of past exposure (28). Although several studies have assessed the effects of vaccination on COVID-19 dynamics, very few have considered its delayed effects.

The aim of this paper is to use an ecological or trend study to evaluate the way how vaccine coverage of the German population is associated with the number of SARS-CoV-2 patients in general hospitals as well as intensive care units. Instead of extrapolating from time trends before vaccine introduction, we adjusted for the remaining time trends by natural splines with a high degree of freedom. We applied a flexible modeling framework by Gasparrini et al. (29) that can simultaneously represent exposure-response dependencies and delayed effects. This family of models is called distributed lag linear models (DLMs). Specifically, we evaluated the effects of vaccination on the number of prevalent hospital patients (hospital cases) and intensive care unit patients (ICU cases) through three separate analyzes by considering people vaccinated

with one dose (i), people vaccinated with two doses (ii), and people vaccinated with three doses (iii). We focused on these two outcomes (hospital cases and ICU cases) since for the COVID-19 pandemic, controlling hospital and ICU admissions was for German public health authorities, an important factor in saving the lives of the patients (30).

2. Methods

2.1. Model framework

2.1.1. General form

To describe the time series of outcomes Y_t, the general form of the model is Gasparrini et al. (29):

$$f\left(E(Y_t)\right) = \alpha + \sum_{j=1}^{J} s_j\left(x_{tj};\beta_j\right) + \sum_{k=1}^{K} \gamma_k u_{tk}, \quad (1)$$

where f is a monotonic link function and Y_t is a count time series response variable, with $t = 1,\ldots,n$, following a distribution that belongs to the exponential family. s_j defines a smoothed relationship between x_j and Y_t through a coefficient β_j. u_k represent confounding variables and γ_k the related coefficients.

In this work, the outcomes Y_t are daily numbers of prevalent hospital patients and intensive care unit patients. According to Cameron and Trivedi test (31), these outcomes are overdispersed (α = 2903.716, $p < 0.0001$ for hospital cases and α = 1201.386, $p < 0.0001$ for ICU cases). We therefore considered a quasi-Poisson model with $\mathrm{E}(Y) = \mu$; $\mathrm{V}(Y) = \phi\mu$, and a canonical log-link in Equation (1). Our motivation to choose the quasi-Poisson model (instead of other alternatives such as the Negative Binomial model) falls in the straightforward interpretation of the results.

2.1.2. Basic functions and delayed effects

The definition of the basis functions relies on two steps. In the first step, the relationship between x_j and $f(E(Y_t))$ is represented by $s(x)$, and is set in Equation (1) as a sum of linear terms (29). This relationship is carried out by the choice of a basis, a space of functions of which s is an element (32). The associated basis functions are some known transformations of the original variable x that generate a new set of variables, termed basis variables (29). Several basis functions have been proposed, and common functions assuming smooth curves, like polynomials or spline functions (33, 34). The basis function is expressed as follows:

$$s\left(x_t;\boldsymbol{\beta}\right) = \mathbf{z}_t^{\top}\cdot\boldsymbol{\beta}, \quad (2)$$

where $z_t^{\top}$ is the tth row of the $n \times v_x$ basis matrix Z. The basis dimension v_x equals the degrees of freedom (df) spent to define the relationship in this space. The unknown parameters are estimated including Z in the design matrix of the model in Equation (1).

In the second step, the delayed effects are considered as an additional dimension. The outcome at a given time t is then explained in terms of past exposures x_{t-l}, where l (the *lag*) represents the elapsed period between exposure and response, here between vaccination and response.

In this study, the maximum lag is fixed at $L = 30$ days, based on previous estimates of the incubation period for COVID-19 (35, 36).

2.1.3. The distributed lag linear models

For a maximum lag L, the additional lag dimension can be expressed by the $n \times (L+1)$ matrix $\mathbf{Q}$, such as:

$$\mathbf{q}_{t.} = \left[x_t,\ldots,x_{t-\ell},\ldots,x_{t-L}\right]^{\top}, \quad (3)$$

with $\mathbf{q}_{t.}$ as the tth row of $\mathbf{Q}$. The vector of lags $\ell = [0,\ldots,\ell,\ldots,L]^{\top}$ corresponds to the scale of the additional dimension. DLMs are specified by the definition of a cross-basis, a bi-dimensional functional space describing at the same time, the shape of the relationship along the predictor x and its distributed lag effects (37). DLMs apply simultaneously the two transformations described in Equations (2), (3). A DLM is expressed by Gasparrini et al. (29):

$$s\left(x_t;\boldsymbol{\eta}\right) = \sum_{j=1}^{v_x}\sum_{k=1}^{v_\ell} \mathbf{r}_{tj}^{\top}\mathbf{c}_{.k}\eta_{jk} = \mathbf{w}_t^{\top}\boldsymbol{\eta}, \quad (4)$$

where $\mathbf{r}_{tj}$ is the vector of lagged exposures for the time t transformed through the basis function j, $\mathbf{C}$ is an $(L+1)\times v_\ell$ matrix of basis variables derived from the application of the specific basis functions to the lag vector $\boldsymbol{\ell}$, the vector $\mathbf{w}_t$ is obtained by applying the $v_x \cdot v_\ell$ cross-basis functions to x_t and $\boldsymbol{\eta}$ a vector of unknown parameters.

2.2. The data

The time series of the COVID-19 data were extracted from the Robert-Koch-Institute website (https://www.rki.de/) and www.corona-datenplattform.de, data platforms for scientific research. The predictors were the daily cumulative proportions of people vaccinated with one dose (V_1), two doses (V_2), three doses (V_3) and the non-pharmaceutical interventions (NPI) index (Figure 1A). The outcomes were the number of prevalent hospital patients (hospital cases) and intensive care unit patients (ICU cases) (Figure 1B). Hospital cases were collected from March 1, 2020, to June 30, 2022, while ICU cases were collected from March 24, 2020, to June 30, 2022.

2.3. Application of DLM to German data

We were interested in the effects of the daily cumulative proportion of people vaccinated with one dose (V_1), two doses (V_2) and three doses (V_3), respectively. As V_1, V_2 and V_3 are highly correlated, we first regressed each series of proportions against the two others and collected the residuals (38), which were considered as a second variable in the DLM. We performed three preliminary separate analyzes as follows:

For the effect of V_1, we considered:

$$model_1 : log(V_1) \sim log(V_2) + log(V_3) + \epsilon_a. \quad (5)$$

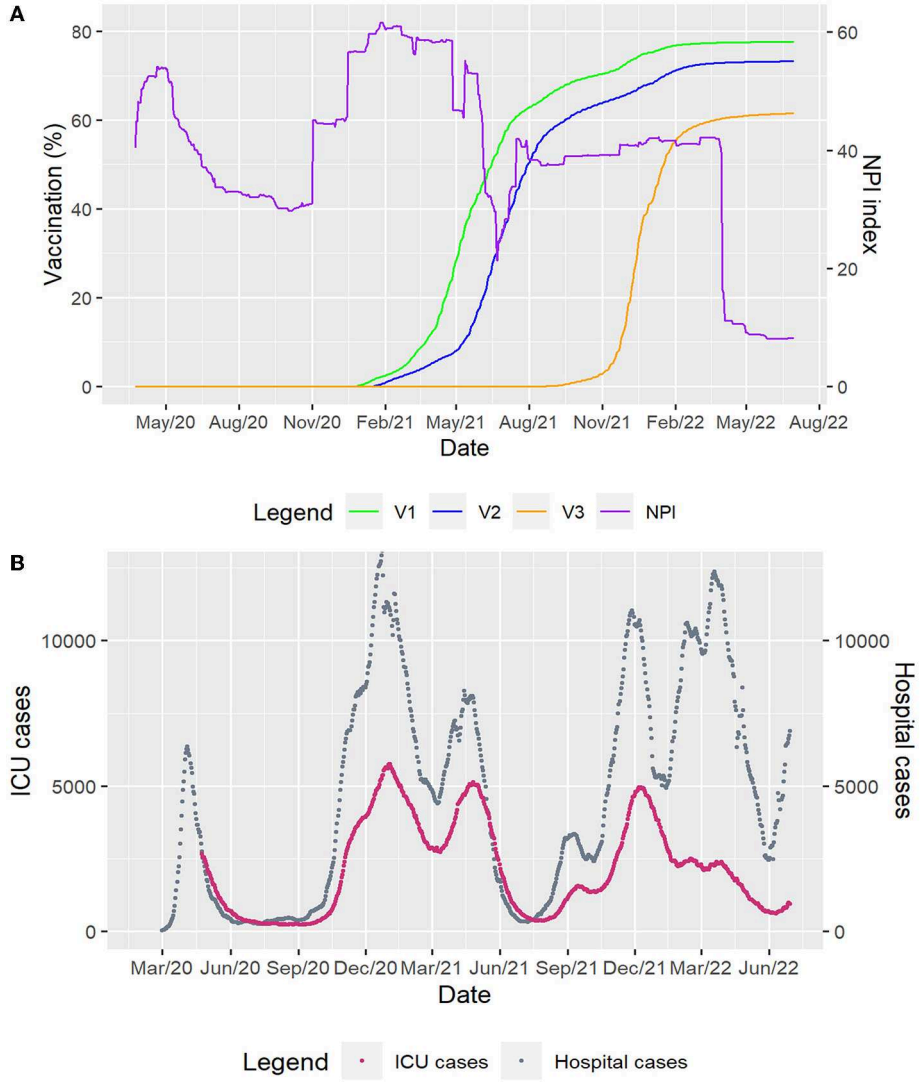


FIGURE 1
Time series of German data considered. **(A)** The predictors represented by the daily cumulative proportions of people vaccinated with one dose (V_1), two doses (V_2), three doses (V_3) and the non-pharmaceutical interventions (*NPI*) index. **(B)** The outcomes: the daily numbers of prevalent hospital patients (hospital cases) and intensive care unit patients (ICU cases).

For the effect of V_2, we used:

$$model_2 : log(V_2) \sim log(V_1) + log(V_3) + \epsilon_b. \quad (6)$$

For the effect of V_3, we applied:

$$model_3 : log(V_3) \sim log(V_1) + log(V_2) + \epsilon_c. \quad (7)$$

The DLMs considered to assess the effects of V_1, V_2, and V_3 on the hospital and ICU cases were defined, respectively, as follows:

$$\log(E(Y_t)) = \alpha_1 + ns(time, df) + \beta_1 V_{1,t} + \gamma_1 \epsilon_{a,t} + \lambda_1 NPI_t, \quad (8)$$

$$\log(E(Y_t)) = \alpha_2 + ns(time, df) + \beta_2 V_{2,t} + \gamma_2 \epsilon_{b,t} + \lambda_2 NPI_t, \quad (9)$$

$$\log(E(Y_t)) = \alpha_3 + ns(time, df) + \beta_3 V_{3,t} + \gamma_3 \epsilon_{c,t} + \lambda_3 NPI_t, \quad (10)$$

where Y_t represents the hospital or ICU cases, log, the natural log function, α_1, α_2, α_3 are the models intercepts, ϵ_a, ϵ_b, and ϵ_c are the residuals extracted from Equations (5)–(7). The variable *time* was set in the model to consider long-term trends and to account for some of the pandemic patterns, such as variants and seasonal variations, which are not explained by remaining predictors. In Equations (8)–(10), $V_{1,t}$, $V_{2,t}$ and $V_{3,t}$ are the cross-basis functions of the three vaccination doses while $\epsilon_{a,t}$, $\epsilon_{b,t}$ and $\epsilon_{c,t}$ represent the cross-basis functions for the residuals and NPI_t, the cross-basis function of the non-pharmaceutical interventions index, considered as confounding variable in the models. The unknown coefficients in the three models are β_1, β_2, β_3, γ_1, γ_2, γ_3, λ_1, λ_2 and λ_3. Moreover, in the Equations (8), (9), and (10), the terms $V_{1,t}$, $V_{2,t}$, $V_{3,t}$, $\epsilon_{a,t}$, $\epsilon_{b,t}$, $\epsilon_{c,t}$ and NPI_t are lagged with lags $\ell \in [0, L]$, where $L = 30$ days represent the maximum lag period. This value

was allocated to the maximum lag period considering previous estimates of the incubation period for COVID-19 (35, 36).

The natural cubic spline *ns*, a flexible and effective technique for adjustment for nonlinear confounding effects (39), was used to adjust for the predictors with two degrees of freedom. This number of degrees of freedom was selected after a sensitivity analysis (40). The natural cubic spline *ns* was also used for the variable *time* and the degrees of freedom (dfs) to find the best modeling of the time trend was chosen by minimizing the Quasi-Akaike information criterion (QAIC) (37) as we considered a quasi-Poisson model framework.

The relative count change (RCC) with a 95% confidence interval (*CI*), calculated as a relative increase/decrease in counts of hospital and ICU patients, was used to assess the effects. RCC = 1 means that there is no connection between vaccination and the disease while RCC $<$ 1 and RCC $>$ 1 are related to the reduction and increase in counts of hospital and ICU patients, respectively (41). Contour plots that depend on lag times and values of V_1, V_2, and V_3 were used to visualize the effects. All analyzes were performed in R version 4.2.1 with the package dlnm (37).

3. Results

To assess the effects of V_1, the best fitting was obtained for 15 and 16 degrees of freedom for ICU and hospital cases, respectively. Concerning V_2, 16 and 20 degrees of freedom for ICU and hospital cases gave the best fitting. Regarding V_3, 16 and 23 degrees of freedom for ICU and hospital cases showed the best fitting. Supplementary Figures S1, S2 show QAIC values and models fitting (observed data and fitted models).

Figure 2 presents the contour plots of the combined effects of lag times and vaccinations on the relative count change (RCC) of hospital and ICU cases. Overall, low vaccine coverage for the first, second and third doses (0–10%) and short (0–4 days) lag times show no connection between vaccination and the number of patients in hospital and ICU (RCC≈1). However, higher vaccine coverages and longer lag times were associated with a bigger decrease in the number of patients in hospitals and ICUs (RCC $<$ 1). The number of COVID-19 patients in hospitals or ICUs significantly decreases as the vaccine coverage increases. Moreover, there were delayed effects of vaccination according to the doses. Strong protective effects were obtained for a lag time of about 15–20 days after vaccination, when at least about 40% of people are vaccinated.

3.1. The effects of V_1 on hospital and ICU cases

Figures 2A, B show the contour plots of the effects of V_1 and lag times on the relative count change (RCC) of hospital and ICU cases. There was no significant effect on hospital and ICU cases (RCC≈1) for low vaccine coverage (0–10%) and short (0–4 days) lag times. Protective effects (decrease in the counts of patients in hospitals and ICUs) were observed around V_1 = 20% with a lag of 5 days, where RCC = 0.80 (95% CI 0.74–0.85) for hospital cases and RCC = 0.92 (95% CI 0.88–0.97) for ICU cases. The number of patients in hospital and ICU (RCC $<$ 1) decreases then sharply as the lag days and vaccine coverage increase. The number of COVID-19 patients in hospitals or ICU significantly decreases (strongest positive effects) for the highest vaccine coverage (77%) and longest lag times (30 days) with RCC = 0.07 (95% CI 0.06–0.09) for hospital cases and RCC = 0.24 (95% CI 0.21–0.27) for ICU cases. Moreover, comparatively, the effects of V_1 on hospital cases are more immediate and intense than on ICU cases.

3.2. The effects of V_2 on hospital and ICU cases

Figures 2C, D show the relative count change (RCC) of hospital and ICU cases as a function of V_2 and lag times. Examining the contour plots, no significant effect of V_2 was observed on hospital and ICU cases (RCC≈1) for low vaccine coverage (0–10%). Considering the hospital cases, moderate and immediate positive effects (decrease in counts of hospital patients) were obtained for moderate vaccine coverage V_2 (20–50%) while strong and immediate positive effects were observed for high vaccine coverage V_2 (50–73%). For ICU cases, there were adverse effects (increase in counts of ICU patients) for the highest vaccine coverage (70%) and short (0–3 days) lag times with RCC=1.04 (95% CI 0.79–1.35). From a lag of 5 days, the highest vaccine coverages V_2 were associated with the lowest RCC values, showing strong and positive effects on ICU cases, which last up to 30 days. The effects of V_2 on hospital cases are more immediate and intense than on ICU cases.

3.3. The effects of V_3 on hospital and ICU cases

Figures 2E, F show the relative count change (RCC) of hospital and ICU cases as a function of V_3 and lag times. No significant effect of V_3 was noticed on hospital and ICU cases (RCC≈1) for low vaccine coverage (0–10%). Regarding the hospital cases, moderate and immediate positive effects (decrease in counts of hospital patients) were obtained for moderate vaccine coverage V_3 (20–35%) while strong and immediate positive effects were observed for high vaccine coverage V_3 (40–61.50%). For ICU cases, there were adverse effects (increase in counts of ICU patients) for the highest vaccine coverage (60%) and short (0–7 days) lag times with RCC = 1.03 (95% CI 1.00–1.07). From a lag of 10 days, the highest vaccine coverages V_3 were associated with the lowest RCC values and strong and positive effects on ICU cases were observed until a lag of 30 days.

Comparison between vaccine coverages shows that V_1 has a more immediate and intense effect than V_2 and that V_2 also has a more immediate and intense effect than V_3. These observations were made for both hospital and ICU cases. Supplementary Figures S3, S4 present RCC point estimates and their confidence intervals for vaccination coverages V1, V2, and V3 in the cases of hospital and ICU patients, respectively.

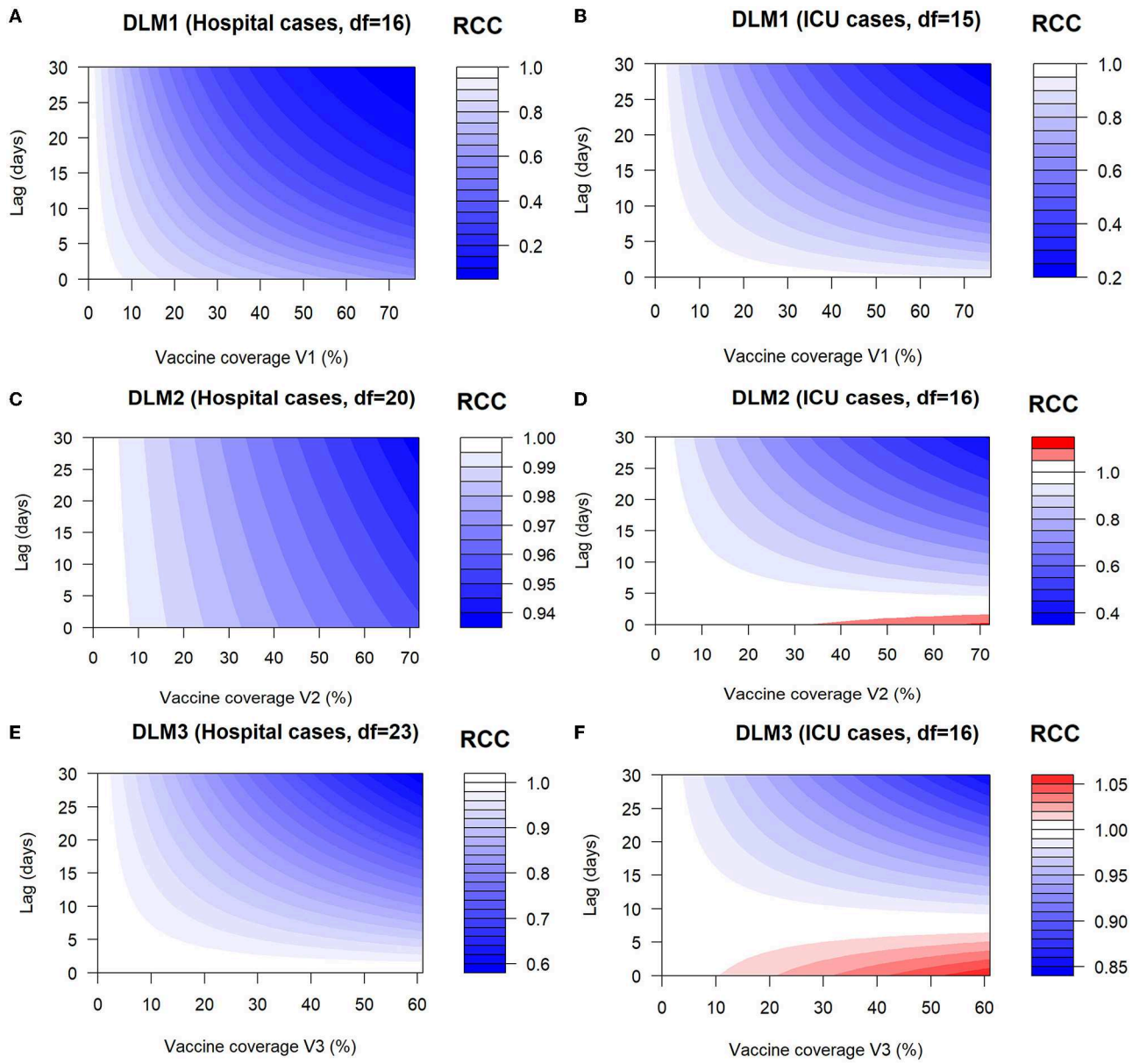


FIGURE 2
Contour plots of the combined effects of lag times and vaccinations on the relative count change (RCC) of hospital and ICU cases. **(A)** Contour plot of RCC of hospital cases as a function of V_1 and lag times. **(B)** Contour plot of RCC of ICU cases as a function of V_1 and lag times. **(C)** Contour plot of RCC of hospital cases as a function of V_2 and lag times. **(D)** Contour plot of RCC of ICU cases as a function of V_2 and lag times. **(E)** Contour plot of RCC of hospital cases as a function of V_3 and lag times. **(F)** Contour plot of RCC of ICU cases as a function of V_3 and lag times. DLM1, DLM2, and DLM3 represent the distributed lag models for the first, second and third doses.

4. Discussion

In this study, we used an ecological or trend study to assess the effects of vaccination and its lag-time on the number of COVID-19 hospitals and ICU patients in Germany. From our results, there was no significant link between the vaccination coverages V_1, V_2, and V_3 and the number of patients in hospital and ICU for low vaccine coverages (0–10%) and short lag times as the relative count change (RCC) was about 1. This means that regardless of the dose of vaccination received, at least 10% of the population must be vaccinated to expect a beginning protective effect against hospital and ICU admissions. The protective effect is low from 10% and then increases as the vaccination rate increases. As expected, this result supports the point that a high vaccination rate is correlated with a lower number of patients in hospital and ICU (20). We also found that, in the context of Germany, the vaccination takes its strong protective effect when at least approximately 40% of people are vaccinated. This result is consistent with those in Springer et al. (22), which show a negative correlation between incidence and vaccination rate in Germany during the 4th wave where the vaccination rate is above 40%. Our findings are also in line with previous studies showing that

COVID-19 vaccines are effective against severe forms of the disease (20, 42).

Furthermore, our results showed a delayed effect of the vaccination according to the doses and outcomes. Indeed, for the hospital cases, the effect is immediate for the first and second doses while for the third dose, a strong protective effect is obtained about 15 days after vaccination. Concerning ICUs cases, there was a lag time of about 15–20 days to obtain a strong protective effect after the first, second and third doses, respectively. These lag times are short compared to that obtained by Li et al. (27) who argued that the lag time for a response to vaccination was at least 40 days. However, contrary to our study, they used the daily reported cases and effective reproduction number as outcomes.

One strength of this study is that our outcomes (hospital and ICU cases) are very specific with low random noise in contrast to other outcomes (general SARS-CoV-2 cases, death cases associated with SARS-CoV-2). German hospitals and intensive care units are legally obligated to report these SARS-CoV-2 data diagnosed with PCR. In Germany, there were about 16.69% and 33.36% deaths among patients admitted to hospital and ICU, respectively (43). These death rates are very high compared to those in the whole population, which is 4.35% (44). It was then important to quantify the effects of vaccination to analyze its contribution to the control of hospital and ICU admissions since public health authorities were most concerned about the scenario where the demand exceeds the capacity of healthcare services (30). To our knowledge, this study is the first that analyzes the effects of vaccination and its lag times on the number of COVID-19 patients in hospitals and ICUs in Germany taking into account long time trends. The findings of this work are relevant and can be applied in other settings and localities. We also included NPIs in our models as a confounding variable since they were maintained at a certain level in the German population while vaccines are distributed. The use of COVID-19 vaccines in combination with the implementation of NPIs is seen as the best alternative to rapidly control the pandemic (45). There is also evidence that an epidemic is likely to rebound immediately after the implementation of a vaccination program if NPIs are completely abandoned (27).

One limitation of our study is that our results are highly dependent on the way we adjusted for the time trends. Moreover, we do not extrapolate hospital cases from time trends before vaccine introduction since emerging variants highly differ from previous variants in terms of transmission, medical condition and burden of disease.

Several observational study designs are discussed in the literature to evaluate the impacts of interventions during an epidemic (13, 46). Digitale et al. (46) reformulated observational studies as pragmatic designs. For these authors, instead of asking retrospective questions about interventions that occurred in the past, the goal should be to prospectively collect data about planned interventions in the future. However, pragmatic designs require more initial planning, community engagement and regulations for human research protection (47). In this study, our design and analysis differ from pragmatic designs. Moreover, we are not interested in estimating vaccine efficacy through randomized trials of individuals or the evaluation of vaccine effectiveness through observational cohort and case-control studies (13). Instead, we aimed to assess the effects of changing vaccine coverage on the number of patients in hospitals and ICUs at a population level. We evaluated the way how vaccination is associated with a decrease in the number of patients in hospitals and ICUs. According to Lipsitch et al. (13), an important consideration of time-trend vaccine effectiveness studies is that the disease outcome under study must be sufficiently specific so that the vaccine's impact on it is likely to be measurable. This was the case for the outcomes considered in this study. Health authorities should therefore consider these results when designing vaccination programs for future pandemics. Knowing the lag times of the vaccination would allow the public health authorities to design appropriate interventions to effectively interrupt the disease transmission (48) if new variants emerged.

5. Conclusion

This work highlights the effects of vaccination on the admission of COVID-19 patients in hospitals and ICUs in Germany. Our results showed a decrease in the number of patients in hospitals and ICUs for an increase in vaccine coverage. This is in line with the protective effects of vaccines against the severe forms of COVID-19 as proved through clinical trials. Moreover, we found that the response to vaccination could be delayed for about 20 days. These findings could be used for designing vaccination programs for future pandemics. Further studies should assess the effects of vaccination considering regional, demographic and social aspects.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.corona-datenplattform.de.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

BEL, RGK, and MW contributed to the study conception and design. BEL carried out data analysis and wrote the initial manuscript draft. MW contributed to data collection, supervised data analysis, and revised the manuscript. RGK supervised data analysis and revised the manuscript. All authors read and revised the initial manuscript. All authors contributed to the article and approved the submitted version.

Funding

The authors acknowledge the support from the Humboldt Research Hub SEMCA funded by the German Federal Foreign Office with the support of the Alexander von Humboldt Foundation (AvH).

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1085991/full#supplementary-material

References

1. Mbow M, Lell B, Jochems SP, Cisse B, Mboup S, Dewals BG, et al. COVID-19 in Africa: dampening the storm? *Science*. (2020) 369:624–6. doi: 10.1126/science.abd3902

2. Maiti A, Zhang Q, Sannigrahi S, Pramanik S, Chakraborti S, Cerda A, et al. Exploring spatiotemporal effects of the driving factors on COVID-19 incidences in the contiguous United States. *Sustain Cities Soc*. (2021) 68:102784. doi: 10.1016/j.scs.2021.102784

3. Ai H, Nie R, Wang X. Evaluation of the effects of meteorological factors on COVID-19 prevalence by the distributed lag nonlinear model. *J Transl Med*. (2022) 20:1–9. doi: 10.1186/s12967-022-03371-1

4. WHO. Weekly Epidemiological Record, 2022. vol. 97, 33 [full issue]. *Weekly Epidemiological Record*. (2022) 97:381–96.

5. Heidrich P, Schäfer M, Nikouei M, Götz T. The COVID-19 outbreak in Germany–Models and Parameter Estimation. *Commun Biomath Sci*. (2020) 3:37–59. doi: 10.5614/cbms.2020.3.1.5

6. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet*. (2020) 395:931–4. doi: 10.1016/S0140-6736(20)30567-5

7. Schäfer M, Wijaya KP, Rockenfeller R, Götz T. The impact of travelling on the COVID-19 infection cases in Germany. *BMC Infectious Dis*. (2022) 22:1–19. doi: 10.1186/s12879-022-07396-1

8. Gumel AB, Iboi EA, Ngonghala CN, Elbasha EH. A primer on using mathematics to understand COVID-19 dynamics: Modeling, analysis and simulations. *Infectious Dis Model*. (2021) 6:148–68. doi: 10.1016/j.idm.2020.11.005

9. Owid. *Coronavirus (COVID-19) Vaccination*. (2022). Available online at: https://ourworldindataorg/covid-vaccinations (assessed on August 14, 2022).

10. Shim E. Projecting the impact of SARS-CoV-2 variants and the vaccination program on the fourth wave of the COVID-19 pandemic in South Korea. *Int J Environ Res Public Health*. (2021) 18:7578. doi: 10.3390/ijerph18147578

11. Choi Y, Kim JS, Kim JE, Choi H, Lee CH. Vaccination prioritization strategies for COVID-19 in Korea: A mathematical modeling approach. *Int J Environ Res Public Health*. (2021) 18:4240. doi: 10.3390/ijerph18084240

12. Olliaro P, Torreele E, Vaillant M. COVID-19 vaccine efficacy and effectiveness–the elephant (not) in the room. *Lancet Microbe*. (2021) 2:e279–80. doi: 10.1016/S2666-5247(21)00069-0

13. Lipsitch M, Jha A, Simonsen L. Observational studies and the difficult quest for causality: lessons from vaccine effectiveness and impact studies. *Int J Epidemiol*. (2016) 45:2060–74. doi: 10.1093/ije/dyw124

14. Pritchard E, Matthews PC, Stoesser N, Eyre DW, Gethings O, Vihta KD, et al. Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom. *Nat Med*. (2021) 27:1370–8. doi: 10.1038/s41591-021-01410-w

15. Mahase E. Covid-19: two vaccine doses are crucial for protection against delta, study finds. *Br Med J Publishing Group*. (2021) 374:n2029. doi: 10.1136/bmj.n2029

16. Horne EM, Hulme WJ, Keogh RH, Palmer TM, Williamson EJ, Parker EP, et al. Waning effectiveness of BNT162b2 and ChAdOx1 covid-19 vaccines over six months since second dose: OpenSAFELY cohort study using linked electronic health records. *BMJ*. (2022) 378:e071249. doi: 10.1101/2022.03.23.22272804

17. Hulme WJ, Williamson EJ, Green AC, Bhaskaran K, McDonald HI, Rentsch CT, et al. Comparative effectiveness of ChAdOx1 versus BNT162b2 covid-19 vaccines in health and social care workers in England: cohort study using OpenSAFELY. *BMJ*. (2022) 378:e068946. doi: 10.1101/2021.10.13.21264937

18. Gnanvi JE, Salako KV, Kotanmi GB, Glèlè KakaïR. On the reliability of predictions on COVID-19 dynamics: a systematic and critical review of modelling techniques. *Infect Dis Model*. (2021) 6:258–72. doi: 10.1016/j.idm.2020.12.008

19. Dashtbali M, Mirzaie M. A compartmental model that predicts the effect of social distancing and vaccination on controlling COVID-19. *Sci Rep*. (2021) 11:1–11. doi: 10.1038/s41598-021-86873-0

20. Wollschläger D, Gianicolo E, Blettner M, Hamann R, Herm-Stapelberg N, Schoeps M. Association of COVID-19 mortality with COVID-19 vaccination rates in Rhineland-Palatinate (Germany) from calendar week 1 to 20 in the year 2021: a registry-based analysis. *Eur J Epidemiol*. (2021) 36:1231–6. doi: 10.1007/s10654-021-00825-6

21. Braun P, Braun J, Woodcock BG. COVID-19: effect-modelling of vaccination in Germany with regard to the mutant strain B. 1.1. 7 and occupancy of ICU facilities. *Int J Clin Pharmacol Therapeut*. (2021) 59:487. doi: 10.5414/CP204064

22. Springer S, Kaatz M, Zieger M. Evaluation of weekly COVID-19 vaccination and case data supports negative correlation between incidence and vaccination in German federal states and cities during 4th wave. *Vaccine*. (2022) 40:2988–92. doi: 10.1016/j.vaccine.2022.04.015

23. Campos M, Sempere JM, Galán JC, Moya A, Cantón R, Llorens C, et al. Simulating the efficacy of vaccines on the epidemiological dynamics of SARS-CoV-2 in a membrane computing model. *microLife*. (2022) 3:18. doi: 10.1093/femsml/uqac018

24. Stiegelmeier EW, Bressan GM, Martinez ALM. The effects of vaccination on COVID-19 dynamics in Brazil: a fuzzy approach. *Braz Arch Biol Technol*. (2022) 65:185. doi: 10.1590/1678-4324-2022220185

25. Sepulveda G, Arenas AJ, González-Parra G. Mathematical Modeling of COVID-19 dynamics under two vaccination doses and delay effects. *Mathematics*. (2023) 11:369. doi: 10.3390/math11020369

26. da Silva CC, de Lima CL, da Silva ACG, Silva EL, Marques GS, de Araújo LJB, et al. Covid-19 dynamic monitoring and real-time spatio-temporal forecasting. *Front Public Health*. (2021) 9:641253. doi: 10.3389/fpubh.2021.641253

27. Li H, Wang L, Zhang M, Lu Y, Wang W. Effects of vaccination and non-pharmaceutical interventions and their lag times on the COVID-19 pandemic: comparison of eight countries. *PLoS Neglected Trop. Dis*. (2022) 16:e0010101. doi: 10.1371/journal.pntd.0010101

28. Gasparrini A. Modeling exposure-lag-response associations with distributed lag non-linear models. *Stat Med*. (2014) 33:881–99. doi: 10.1002/sim.5963

29. Gasparrini A, Armstrong B, Kenward MG. Distributed lag non-linear models. *Stat Med*. (2010) 29:2224–34. doi: 10.1002/sim.3940

30. Chadsuthi S, Modchang C. Modelling the effectiveness of intervention strategies to control COVID-19 outbreaks and estimating healthcare demand in Germany. *Public Health Practice*. (2021) 2:100121. doi: 10.1016/j.puhip.2021.100121

31. Cameron AC, Trivedi PK. Regression-based tests for overdispersion in the Poisson model. *J Econometr*. (1990) 46:347–364. doi: 10.1016/0304-4076(90)90014-K

32. Wood SN. *Generalized Additive Models: An Introduction With R*. New York, NY: Chapman and Hall/CRC.

33. Braga ALF, Zanobetti A, Schwartz J. The time course of weather-related deaths. *Epidemiology*. (2001) 12:662–7. doi: 10.1097/00001648-200111000-00014

34. Dominici F. Time-series analysis of air pollution and mortality: a statistical review. *Res Rep Health Eff Inst*. (2004) 123:3–27.

35. Yin MZ, Zhu QW, Lü X. Parameter estimation of the incubation period of COVID-19 based on the doubly interval-censored data model. *Nonlinear Dyn*. (2021) 106:1347–58. doi: 10.1007/s11071-021-06587-w

36. Bikbov B, Bikbov A. Maximum incubation period for COVID-19 infection: do we need to rethink the 14-day quarantine policy? *Travel Med Infect Dis*. (2021) 40:101976. doi: 10.1016/j.tmaid.2021.101976

37. Gasparrini A. Distributed lag linear and non-linear models in R: the package dlnm. *J Stat Software*. (2011) 43:1. doi: 10.18637/jss.v043.i08

38. García CB, Salmerón R, García C, García J. Residualization: justification, properties and application. *J Appl Stat*. (2020) 47:1990–2010. doi: 10.1080/02664763.2019.1701638

39. Xu Q, Li R, Rutherford S, Luo C, Liu Y, Wang Z, et al. Using a distributed lag non-linear model to identify impact of temperature variables on haemorrhagic fever with renal syndrome in Shandong Province. *Epidemiol Infect*. (2018) 146:1671–9. doi: 10.1017/S095026881800184X

40. Runkle JD, Sugg MM, Leeper RD, Rao Y, Matthews JL, Rennie JJ. Short-term effects of specific humidity and temperature on COVID-19 morbidity in select US cities. *Sci Total Environ*. (2020) 740:140093. doi: 10.1016/j.scitotenv.2020.140093

41. Wong TW, Lau TS, Yu TS, Neller A, Wong SL, Tam W, et al. Air pollution and hospital admissions for respiratory and cardiovascular diseases in Hong Kong. *Occupat Environ Med*. (1999) 56:679–683. doi: 10.1136/oem.56.10.679

42. Moghadas SM, Vilches TN, Zhang K, Wells CR, Shoukat A, Singer BH, et al. The impact of vaccination on coronavirus disease 2019 (COVID-19) outbreaks in the United States. *Clin Infect Dis*. (2021) 73:2257–64. doi: 10.1093/cid/ciab079

43. Kloka JA, Blum LV, Old O, Zacharowski K, Friedrichson B. Characteristics and mortality of 561,379 hospitalized COVID-19 patients in Germany until December 2021 based on real-life data. *Sci Rep*. (2022) 12:1–9. doi: 10.1038/s41598-022-15287-3

44. Bhapkar H, Mahalle PN, Dey N, Santosh K. Revisited COVID-19 mortality and recovery rates: are we missing recovery time period? *J Med Syst*. (2020) 44:1–5. doi: 10.1007/s10916-020-01668-6

45. Rowan NJ, Moral RA. Disposable face masks and reusable face coverings as non-pharmaceutical interventions (NPIs) to prevent transmission of SARS-CoV-2 variants that cause coronavirus disease (COVID-19): Role of new sustainable NPI design innovations and predictive mathematical modelling. *Sci Total Environ*. (2021) 772:145530. doi: 10.1016/j.scitotenv.2021.145530

46. Digitale JC, Stojanovski K, McCulloch CE, Handley MA. Study designs to assess real-world interventions to prevent COVID-19. *Front Public Health*. (2021) 9:1063. doi: 10.3389/fpubh.2021.657976

47. Casey JD, Beskow LM, Brown J, Brown SM, Gayat É, Gong MN, et al. Use of pragmatic and explanatory trial designs in acute care research: lessons from COVID-19. *Lancet Respiratory Med*. (2022) 10:700–14. doi: 10.1016/S2213-2600(22)00044-3

48. Okiring J, Routledge I, Epstein A, Namuganga JF, Kamya EV, Obeng-Amoako GO, et al. Associations between environmental covariates and temporal changes in malaria incidence in high transmission settings of Uganda: a distributed lag nonlinear analysis. *BMC Public Health*. (2021) 21:1–11. doi: 10.1186/s12889-021-11949-5

frontiers | Frontiers in Public Health

TYPE Review
PUBLISHED 11 April 2023
DOI 10.3389/fpubh.2023.1140353

OPEN ACCESS

EDITED BY
Pierpaolo Ferrante,
National Institute for Insurance Against Accidents at Work (INAIL), Italy

REVIEWED BY
Ioannis Kokkinakis,
University Center of General Medicine and Public Health, Switzerland
Adelia Sequeira,
University of Lisbon, Portugal

*CORRESPONDENCE
Marcelo A. Navarrete
✉ marcelo.navarrete@umag.cl

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Infectious Diseases: Epidemiology and Prevention, a section of the journal Frontiers in Public Health

RECEIVED 08 January 2023
ACCEPTED 20 March 2023
PUBLISHED 11 April 2023

CITATION
Sarmiento Varón L, González-Puelma J, Medina-Ortiz D, Aldridge J, Alvarez-Saravia D, Uribe-Paredes R and Navarrete MA (2023) The role of machine learning in health policies during the COVID-19 pandemic and in long COVID management. *Front. Public Health* 11:1140353. doi: 10.3389/fpubh.2023.1140353

The role of machine learning in health policies during the COVID-19 pandemic and in long COVID management

Lindybeth Sarmiento Varón[1†], Jorge González-Puelma[1,2†], David Medina-Ortiz[3], Jacqueline Aldridge[3], Diego Alvarez-Saravia[1,2], Roberto Uribe-Paredes[3] and Marcelo A. Navarrete[1,2*]

[1]Centro Asistencial Docente y de Investigación, Universidad de Magallanes, Punta Arenas, Chile, [2]Escuela de Medicina, Universidad de Magallanes, Punta Arenas, Chile, [3]Departamento de Ingeniería en Computación, Facultad de Ingeniería, Universidad de Magallanes, Punta Arenas, Chile

The ongoing COVID-19 pandemic is arguably one of the most challenging health crises in modern times. The development of effective strategies to control the spread of SARS-CoV-2 were major goals for governments and policy makers. Mathematical modeling and machine learning emerged as potent tools to guide and optimize the different control measures. This review briefly summarizes the SARS-CoV-2 pandemic evolution during the first 3 years. It details the main public health challenges focusing on the contribution of mathematical modeling to design and guide government action plans and spread mitigation interventions of SARS-CoV-2. Next describes the application of machine learning methods in a series of study cases, including COVID-19 clinical diagnosis, the analysis of epidemiological variables, and drug discovery by protein engineering techniques. Lastly, it explores the use of machine learning tools for investigating long COVID, by identifying patterns and relationships of symptoms, predicting risk indicators, and enabling early evaluation of COVID-19 sequelae.

KEYWORDS

COVID-19, public health policies, mathematical models, machine learning, long COVID, SARS-CoV-2

1. Introduction

Mathematical models help to understand the functioning and dynamics of a given system trough equations and rules, as such, can simulate conditions and scenarios associated with multiple public policies, non-pharmaceutical interventions (NPI), and vaccine performance (1). Therefore, mathematical models became major tools for guiding the decision-making of governments and health systems during the pandemic (2). This section briefly introduces SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) and describes relevant events during the progress pandemic. We then summarize the main applications of mathematical models and the various uses to describe the transmission behaviors of SARS-CoV-2.

1.1. What is the SARS-CoV-2?

SARS-CoV-2 is the pathogen causing the 2019 coronavirus disease (COVID-19). COVID-19 manifestations range from mild flu symptoms to severe acute respiratory syndrome. SARS-CoV-2 virion contains a 29Kb RNA genome wrapped in a capsid covered by the Spike, the main protein responsible for the high infection rate (3). During the transmission between humans, the genome accumulates mutations, generating variants with selective advantages that predominate in different countries (4). Intrinsic factors like transmissibility and natural mutation rate, host factors such as age, risk group, immunity, and socio-cultural factors like economy, culture, and current levels of globalization have determined the coronavirus evolution. Integrating SARS-CoV-2 data is essential to predict its behavior, prevent its continuous expansion, and understand this disease. Three years after the pandemic, the scientific community has generated an unprecedented amount of data, now facing the challenge of translating this data into knowledge.

1.2. The first 3 years of the pandemic

The first reported case of COVID-19 was in December 2019 (5). With the exponential increase in infections worldwide, the World Health Organization (WHO) declared the disease a pandemic in March 2020. Governments adopted different NPI to mitigate the virus's high reproduction rates. These measures included face masks, social distancing, and lockdowns. While these measures were implemented worldwide, just a few countries, such as Vietnam and New Zealand, demonstrated the complete -although transitory- elimination of the transmission (6). In April and May 2020, the first predictions of the pandemic course were based on statistical models performed by the Institute for Health Metrics and Evaluation and provided a reasonable projection in the short-time (7). During the first wave, it was also possible to establish that 10% of the cases were responsible for 80% of the secondary infections, indicating a high heterogeneity in transmission spread as compared to other pathogens (8).

In the first pandemic year, it was identified that social contact in public transport or closed areas allowed high transmission rates (9, 10). In turn, it was determined that face masks reduce droplet particle transmission (11). Furthermore, NPI was essential to flatten the spread curve in the first year of the pandemic preventing new waves of cases after curves pick, limiting overcrowding of hospital beds, and giving time to improve treatment strategies (12). Adaptations of the Susceptible-Infected-Recovered models helped to demonstrate the NPI effectiveness in preventing the transmission of the virus. Besides, these same models allowed the detection of an increase in virus circulation with the relaxation of the measures (13). Other models, facilitated the test-track-isolation developing strategies to prevent the spread, demonstrating that efficient track strategies help to reduce the number of new cases (14). At the same time, the first signs of SARS-CoV-2 genetic adaptation arose between March and May 2020, with the emerging D614G variant, which showed clear worldwide transmissibility advantages (15). The control of the pandemic at that time relied on the development of herd immunity, being established that the necessary protection of the population is approximately $1 - \frac{1}{R_0}$, being estimated at 67% of the people (16). In August 2020, reinfection cases demonstrated that natural immunity only provides temporary protection (17). In December 2020, the first clinical trials of vaccines were developed, leading to the emergency approval of traditional and novel vaccine formulations -such as mRNA vaccines-. These studies quickly established that immunity begins between 10 and 14 days after the first dose (18). A second dose shows protection over 90%, preventing hospitalizations and deaths (19). The vaccines can block propagation, making cases less infectious, with a 92% reduction in transmission rates (20). At the same time, quantitative models pointed to the possibility of immune escape when complete schemes are not generated. At the end of 2020, the Alpha variant (B.1.1.7), according to the WHO terminology, was the variant responsible for the significant increase in cases in the United Kingdom. This variant was characterized by presenting spike mutations with binding advantages to the ACE-2 receptor (21), showing clear selection advantages, a phenomenon observed simultaneously in different parts of the world (22, 23).

The subsequent variant of similar global relevance was Delta (B.1.617.2), characterized by its high replicative capacity. Vaccine effectiveness studies showed protection against Alpha and Delta variants (24). Vaccination programs were effective reducing deaths, hospitalization admission, and intensive care unit (ICU) occupancy (see Figure 1). In November 2021, a new outbreak was reported in South Africa, caused by a new circulating variant presenting a 60–70 spike gene deletion. This variant was called Omicron (B1.1.529) and expanded rapidly throughout the world, replacing the Delta variant. Omicron carries more than 30 spike mutations (25), being responsible for high worldwide reinfection rates (26). Vaccines have also shown a protective effect against this variant, although deaths were reported among unvaccinated individuals. Omicron subvariant (XBB1.5) has been described as responsible for 40.5% confirmed cases in the EE.UU. as of late December 2022. It has also been observed that recombinant XBB and BA.2 Omicron subvariant strains, widely spread in Asia, do not show different symptoms than the previous variants, nor do they show signs of being more severe than their predecessors.

Figure 1 summarizes the key variables depicting the pandemic evolution in five exemplary cases. Each country showed different spread behaviors of SARS-CoV-2. The measures showed variable effectiveness. In most countries, other public health policies and government plans were applied to mitigate the effects of the spread. However, in most cases, the fatalities decreased after implementation.

1.3. Applications of mathematical modeling during the pandemic

The SIR models (Susceptible, Infected, and Recovered) are spread dynamics analysis models used during the early days of the pandemic (30). SEIR models (Susceptible-Exposed-Infected-Recovered) correspond to an adapted SIR model to understand propagation mechanisms (31). These models do not account for heterogeneity within the population, thus novel strategies

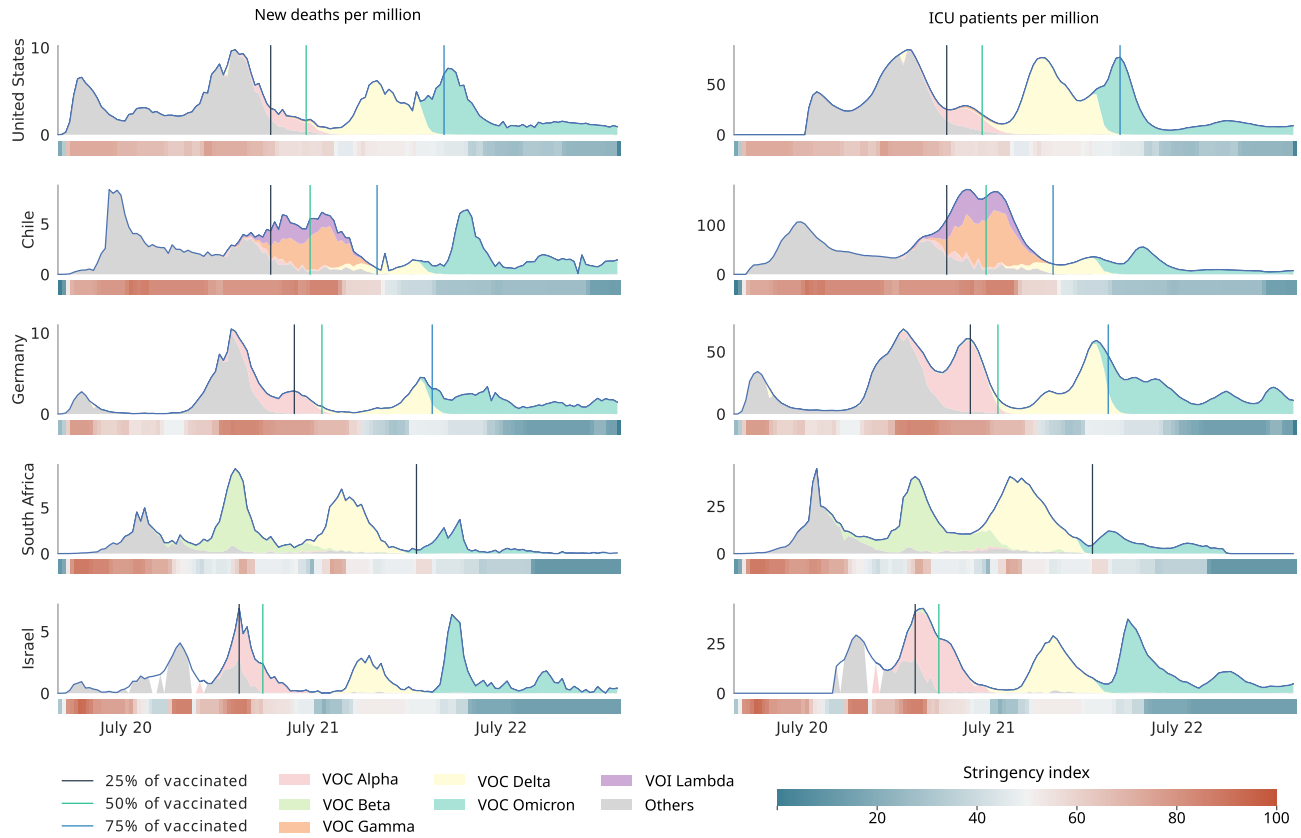


FIGURE 1
Behavior of epidemiological variables during the ongoing pandemic of COVID-19. The figure depicts the timeline of new deaths per million inhabitants **(left)** and admissions to intensive care unit (ICU) per million inhabitants **(right)** in relationship with SARS-CoV-2 variants, vaccination thresholds, and non-pharmacological interventions. The stringency index is a composite measure based on nine response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest). Data acquired from et al. (27), Hasell et al. (28), and Khare et al. (29).

incorporated a component of population subdivision into multiple groups and interconnected systems, allowing the representation of several mechanisms of interaction between different sub-populations by a multi-group SEIRA (Susceptible-Exposed-Infected-Recovered and Asymptomatic Model) (32). Another interesting development was the statistically-based temporal reclassification of cases. This approach allowed more precise modeling of SARS-CoV-2 propagation dynamics, by correcting errors in diagnostic test reporting times and infection time registries (33, 34).

With the application of NPI strategies to prevent the spread of SARS-CoV-2, the mathematical models were adapted to incorporate this new knowledge. This adaptations enabled the anticipation of the effect of NPI relaxation measures in function of epidemiological variables, such as levels of hospitalization, use of ICU, and lethality (35). SEIRA models also helped to asses the effect of vaccines and pharmaceutical interventions (36).

With the first vaccination plans and high immunization rates started the relaxation of public policies (37). However, the ability of the virus to mutate and generate variants was associated with new peaks in cases incidence. Mathematical models were adapted to this scenario by incorporating information on genomic surveillance programs, spread of variants, and the effects of immunization (38–40).

Altogether, mathematical tools proved its relevance in modeling the behavior of propagation systems and their effect on populations. The SIR classical model as well as different adaptations such as SEIR, SEIRA, and others, contributed significantly to the development of government plans and public health policies. Nevertheless, traditional mathematical modeling strategies rely on existing knowledge and cannot account for dynamics not explicitly incorporated during modeling. Methods based on machine learning (ML) and artificial intelligence (AI) can overcome these intrinsic limitations by generating autonomous systems that learn from the modeled dynamics to predict new behaviors and adapt to unknown scenarios.

1.4. Vaccines developments, efficacy, and adverse effects

Population immunity is considered a landmark for epidemic control. Since immunity through natural infection might result in unacceptable morbidity and mortality, the development of efficient

COVID-19 vaccination programs was a prioritary public policy for most countries (41, 42). The race to develop highly effective and safe vaccines resulted in various platforms allowing their implementation at unprecedented speed (43–45).

Due the modest response of traditional vaccines against other coronaviruses such as Middle East Respiratory Syndrome Coronavirus (MERS) and Severe Acute Respiratory Syndrome (SARS), the development of novel formulations was a major scientific goal (42, 46). A new vaccine technology based on mRNA technology emerged as candidates in late December 2020 and two formulations granted emergency approval BNT162b2 (Pfizer-BioNTech), and mRNA-1273 (Moderna) (47). The developed vaccines showed promising results in reducing transmissibility and the probability of death, reaching an efficacy > 90% in phase III clinical trials (48).

The widespread immunization poses the challenge of quantifying and understanding short- and long-term toxicity for novel vaccine formulations. Most studies have shown short-term safety in the general population. However, in certain groups, severe adverse events were reported i) anaphylaxis (2.5–4.8 cases per million adult vaccine doses administered) (49, 50), ii) myocarditis (52.4 cases and 56.3 cases per million doses) (51), iii) thrombosis with thrombocytopenia syndrome (2-4 cases per one million doses administered) (52), and iv) Guillain-Barré syndrome (7.8 cases per million) (53), as well as an association with multisystemic inflammatory syndrome (54).

A major challenge is to reliably detect long-term effects that might occur at different rates in different patients subgroups (55). Causal association becomes difficult due to the high immunization rates achieved in most countries. In this complex scenario mathematical models, ML, and AI, could provide powerful tools provided that public policies focus on collection of sufficient high-quality data.

1.5. What is long COVID?

1.5.1. Characteristics and definitions of long COVID

Long COVID (LC) is a novel multi-systemic disease defined by the persistence or appearance of a wide variety of symptoms with variable intensity, regardless of the initial disease severity by probable or confirmed SARS-CoV-2 infection (56). In response to the absence of a consensus definition, the WHO proposed using the term Post-COVID-19 listed in the ICD-10 classification based on the Delphi consensus (57). This condition usually manifests 3 months after the SARS-CoV-2 infection, the symptoms last for at least 2 months in the absence of alternative diagnosis (58).

The National Institute for Health Research, classifies LC into i) post-intensive care syndrome (post-ICU syndrome), ii) post-viral fatigue syndrome, iii) permanent organ damage, iv) decompensation of previous chronic diseases, v) the onset of a new disease triggered by COVID-19, and vi) pharmacological toxicity from COVID-19 treatment (59).

Other authors had suggested six post-COVID syndrome subsets, including i) non-severe COVID-19 multiorgan sequelae, ii) pulmonary fibrosis sequelae, iii) myalgic encephalomyelitis/chronic fatigue syndrome, iv) postural orthostatic tachycardia syndrome, v) post-intensive care syndrome, and vi) medical or clinical sequelae (60).

1.5.2. Symptoms and incidence of long COVID

Between 2.3 and 60% of COVID-19 survivors could experience LC symptoms during the first year, and up to 42% 2 years after the infection (61–63). Patients with LC present variable symptoms, including fatigue (29%), muscle pain, palpitations, cognitive impairment (28%), dyspnea (21%), anxiety (27%), chest pain, and arthralgia (18%) (see Figure 2) (64). Other patients report respiratory system dysfunction (26%), or cardiovascular complications (32–89%) 3 months after the onset of infection (65–67). Gastrointestinal symptoms have been associated with an imbalance of gut microbiota, as well as psychological and central nervous system effects (68, 69). Most of these symptoms are associated with a reduction in the quality of life. However, the distinction between SARS-CoV-2-related symptoms to those linked to other, often pre-existing conditions remains extremely challenging. As clinical studies addressing this issues take a long time to develop the NIH launched the Rapid Acceleration of Diagnostics initiative, and the NIH LC Computational Challenge (70). This initiative aims to use AI and ML to predict which patients with SARS-CoV-2 infections are most likely to develop LC. Figure 2 depicts the relative frequency of LC symptoms registered by the National COVID Cohort Collaborative (N3C) initiative. Inviduals that tested positive for SARS-CovV-2 show a higher frequency of alterations in symptoms such as fatigue and shortness of breath. The prevalence of these symptoms seems higher in women. However, the small magnitude of the differences highlights the challenge of differentiation long COVID from other conditions.

2. Machine learning application to COVID-19

During the COVID-19 pandemic, ML methods have played a relevant role in the development of diagnostic strategies (72, 73), forecasting the epidemiological behavior (74), and as a tool to support the development and monitoring of public health policies (75). Figure 3 summarizes the most relevant ML applications during the COVID-19 pandemic.

2.1. COVID-19 diagnosis

Different strategies based on ML algorithms were designed during the COVID-19 pandemic to elaborate predictive models of efficient clinical diagnosis (76). The main inputs used to build the models are based on images, sounds, respiratory information, symptoms, and mixed data (77). Convolutional neural networks (CNN) architectures are commonly employed to develop classification models *via* image inputs (e.g., x-ray, CT-chest, and ultrasounds) (78). Sounds from respiratory information, such as cough and breath, were common inputs for the development of

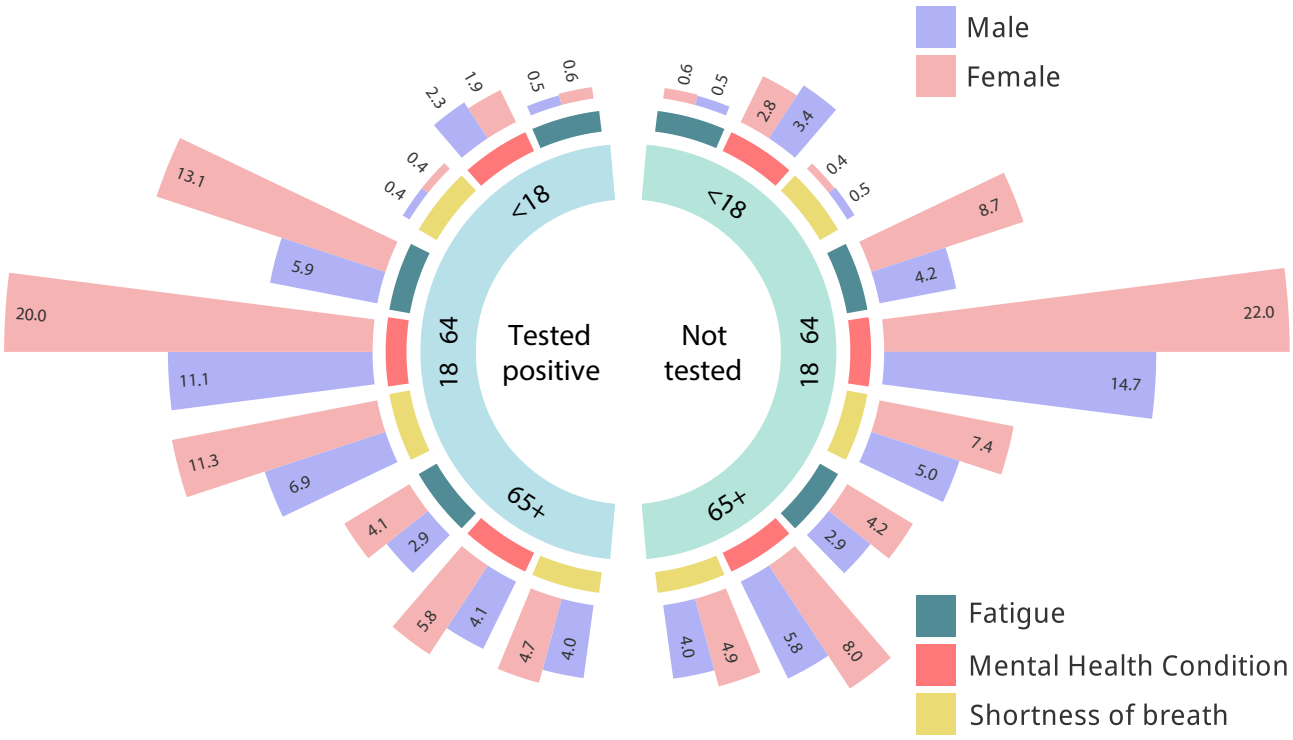


FIGURE 2
Analysis of long COVID symptoms in patients with positive or negative COVID-19 PCR test. Relative frequency of symptoms in individuals with a positive COVID-19 PCR test (left) as compared to individuals with a negative test. Elaborated on basis of LC symptoms registered by the National COVID Cohort Collaborative (N3C) initiative. Data acquired from (71).

predictive models employing recurrent neural network (RNN) or long short-term memory architectures (LSTM), since this type of architectures have the advantage to maintain the information on signal frequencies (79). Hybrid methods that combine symptoms and clinical diagnostic tests as inputs facilitate the development of more complex predictions models or classifications systems. The hybrid methods include not only vector information or matrix spaces, but also data on disease's propagation. The incorporation of virus characteristics, close contacts, and contagion networks using graph neural networks results in highly efficient prediction systems (80).

To demonstrate the usability of classification models based on ML techniques, a clinical diagnostic model using CT chest images was developed following the architecture proposed in Figure 4 and updating our previously reported method for CT chest images classification (34). Generally, models based on CNN architectures can be divided into three large blocks: i) pattern processing and extraction, ii) learning, and iii) classification blocks. To extract patterns, a set of three layers composed of CNN, batch normalization, max pooling, and dropout, was developed. Then, a flattened layer is used to prepare the inputs to the fully connected or dense layers, which are part of the learning block, composed of dense layers interspersed with batch normalization, ending with a dropout layer. Finally, a last layer of classification is added to develop the outputs. As activation functions, ReLU and SoftMax were used. In addition, binary cross entropy associated with an Adam optimizer was used as a cost function. A total of 2,482 images were used to train the diagnostic model extracted from (81). For the training process, a classic validation approach was followed by segmentation of the training and validation data set (80:20), and the TensorFlow framework was employed for its implementation (82). Model training was followed for a total of 10 epochs. The proposed architecture achieved a precision of 99.81% and 0.027 loss function, demonstrating the high performance obtained by the proposed architecture. The implemented model can be used as a support strategy for clinical diagnosis in patients with COVID-19. Besides, it is possible to apply transfer learning techniques to use the same images and the same architecture proposed to estimate the probability that patients present sequelae, one of the most recent areas of study associated with the concept of LC.

2.2. COVID-19 treatments and strategies to prevent adverse effects

ML applications related to the design of treatment strategies have focused on drug discovery, drug repurposing, and vaccine discovery methods (83). For drug repurposing, algorithms are usually based on networks of knowledge graphs including virus and host interactions (84). These strategies have used

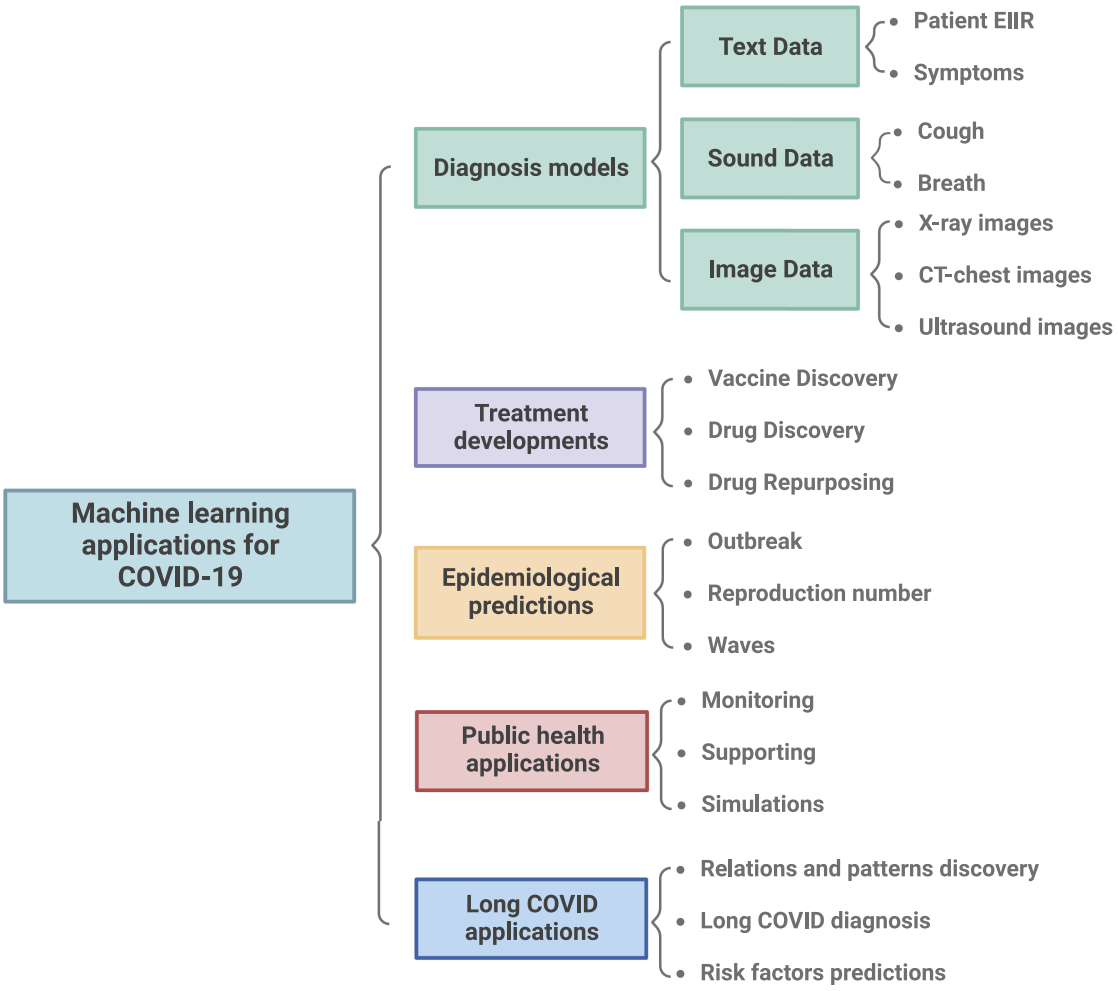


FIGURE 3
Summary of machine learning applications to fight COVID-19 during the pandemic. General applications of machine learning were classified into 5 categories: i) The design of diagnosis models based on different types of inputs like CT chest, X-ray images, and symptom descriptions. ii) Treatment development. iii) The development of epidemiological models to predict new waves and outbreaks. iv) The simulation of potential scenarios, and monitoring systems to guide public health decisions. v) The diagnosis and identification of risk factors in long COVID.

particular network label propagation combined with semi-supervised learning method based on regularized Laplacian to identify interactors of SARS-CoV-2 (85). Another example is the elaboration of predictive systems based on protein-protein interaction to estimate affinity between two elements (86). This issue has been addressed by either CNN or graph convolutional neural networks (GCNN) architectures. Protein complexes are typically represented using strategies based on topological information (87), solvent accessible surface (SAS) (88), voxel-based molecular surface representation (89), and various molecular descriptors (90).

Another of the traditional drug repurposing methods are the gene expression based algorithms (83). The changes in the expression levels of defensive genes in disease states can be used as phenotypic descriptors or quantifiers of the transcriptomic effects of the explored drugs. Besides, methods based on integrated docking simulation algorithms have made it possible to optimize drug repurposing systems (91).

Different computational tools have been developed for drug and vaccine discovery. Zhavoronkov et al. (92) developed a generative chemistry pipeline based on the knowledge of protein, molecule structures, and homology models strategies to identify new drugs related to SARS-CoV-2. Tang et al. (93) have built processes based on deep learning (DL) algorithms to design new antivirus drugs of a chemical or peptide nature based on the information available in the literature and different chemical rules.

Molecular simulations using docking techniques allowed the development of virtual screening methodologies and iterative searches to discover new drugs of interest. The discovery of new chemical compounds with desirable activities is possible by combining the structural information with strategies of deep generative models (94, 95).

Predictive models using the linear protein sequences and the chemical compounds represented as SMILES have been proposed to predict affinity between proteins and chemical compounds (96). Different numerical representations strategies have been

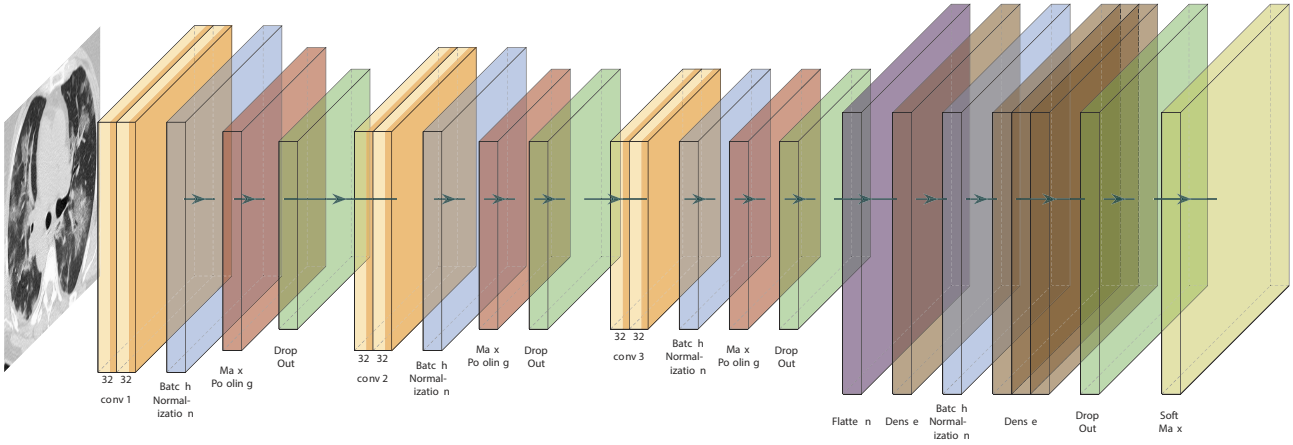


FIGURE 4
Developed architecture for COVID-19 diagnosis classification models based on CT chest images and convolutional neural network architectures. Three blocks of layers composed of convolution, batch normalization, max pooling and dropout layers are generated as a pattern extraction strategy, then a flatten layer is used to generate the inputs to the dense layers, which are joined with a layer of batch normalization, followed by three additional full connected layers, which end with a new dropout layer to prevent overfitting, and the final classification layer. ReLU is used as activation functions and the SoftMax function in the classification layer. Finally, the Adam optimizer is used as a loss function binary cross entropy. The developed architecture is an update from previous method for CT chest images classification models developed by our group (34).

implemented to encode the protein sequences, such as binarization coding, physicochemical properties, and Fourier transforms to represent protein sequences in spaces of signals (97). Alternatively, methods based on natural language processing (protein language models) have been developed (98). In the case of SMILES, different autoencoders and transformers strategies have been created, including variational autoencoders and graph junction trees (99).

Performance between methods based on linear sequences information and those that only incorporate structural details are similar. However, the processes that use representations based on NLP seem to present a higher performance because the autoencoders manage to learn the structural relationships that guide the function (100). Nevertheless, the learning strategies and the abilities to extract complex patterns from the information used for the development of predictive models are properties of DL methods that, to date, have not been fully understood due to their functioning as black boxes. The incorporation of techniques based on explainable AI, is under development to understand the underlying functions and mechanics of the ML algorithms (101).

Concerning the strategies to prevent the adverse effect provoked by the vaccination programs, ML analyzes revealed distinct arterial pulse variability according to side effects of mRNA vaccine. This can facilitate a time-saving and easy-to-use method for detecting changes in the vascular properties associated with cardiovascular side effects following vaccination (102).

The application of explainable ML techniques has allowed to detect relevant variables to perform predictive models with hight performances. Abbaspour et al. (103) applied SHAP strategies combined with XGB model to identify important predictors (e.g., demographics, any history of allergy, any prior COVID diagnosis or positive test, vaccine manufacturer, and time-of-day-of-vaccination) associated to COVID-19 vaccine-related side effects.

Analyzes of the Vaccine Adverse Event Reporting System datasets with ML and a statistical approaches identified and classified pre-existing factors as having an impact on post-vaccination morbidity and reactogenicity (104). Nevertheless, this information is limited because the main databases do not have a larger record size and do not cover all types of vaccines, provoking problems in the generalization of the identified behaivors.

2.3. COVID-19 epidemiology

The design and implementation of ML models used for predicting epidemiological variables was a significant challenge. The need of high volumes of data to generalize the behavior of the predictive models (105), made necessary to develop methods for optimizing the representation of the inputs by autoencoders or embedding (106). The developed models were generated to promote the implementation of computer systems for the simulation of scenarios (107) and to facilitate the elaboration of government public policies focused on preventing the increase in the number of contagious or the outbreak of new waves (108).

Depending on the input type, the construction of predictive models can be based on forecasting methods using strategies such as ARIMA (109, 110) or LSTM architectures (111). Other strategies were based on logistic regression methods (112), nonlinear regressions (113), autoregressive models (114), and Gaussian Process Regression (115). The inputs used to develop the predictive models contemplate information based on time series and consider contagion spread records, NPI, scenarios, and different types of crucial information related to epidemiological variables. Mathematical methods based on linear algebra and kernel applications were used to combine the different kinds of data in hybrid systems elaborated with RNN and CNN architectures (116).

2.4. COVID-19 public health

One essential use of ML strategies was combining mathematical models to develop hybrid knowledge systems to support decisions in public health. These systems can be classified mainly into monitoring applications and simulation systems (116). Concerning monitoring tools, predictive models allow the generation of early alerts of behaviors during a pandemic. These alerts were usually related to predicting waves and new contagion outbreaks. More limited strategies but with significant impact were the methodologies to forecast the level of ICU occupancy in hospitals and health systems and their correlation with increases in contagion rates and mutational variants since it allowed early warning of the occupancy level and facilitated decision-making to prevent a whole occupancy level (117).

The simulation of scenarios by ML allowed the evaluation of public policy effect on populations of interest (118). Despite the versatility of ML, dynamic changes in the knowledge embedded in the system—NPI modifications, the application of vaccine programs, emergence of SARS-CoV-2 variants, etc- makes necessary a constant adaptation of ML based models. Incorporation of reinforced learning might help to facilitate this process.

2.5. Application to long COVID

With the emergence of LC, ML methods have been employed for the development of predictive tools, the construction of statistical systems for relating patient phenotypes, and the elaboration of rules and complex patterns to understand the interactions between systems and types of sequelae. The application of unsupervised learning algorithms like *k*-means and kernel representations strategies enabled to correlate symptoms and different classifications of LC (119).

Based on data from the N3C electronic health record repository, Pfaff et al. (119) have developed an ML model to classify the likelihood of LC diagnosis. Using XGBoost machine learning algorithm this study identified a series of features, including the healthcare utilization rate, patient age, dyspnea or respiratory symptoms, other pre-existing risk factors (diabetes, kidney disease, congestive heart failure, or pulmonary disease), and treatment medication information to predict LC.

Binka et al. (120) proposed a classification model based on elastic net penalized logistic regression algorithms for classifying patients as positive or negative for LC. The model proposed by Binka et al. (120) employed as descriptors demographic characteristics, pre-existing conditions, COVID-19 related data, and all symptoms/conditions recorded >28–183 days after the COVID-19 symptom onset/reported.

Fritsche et al. (121) described associations from the previous and acute medical phenomena of COVID-19 as predisposing diagnoses for LC employing statistical and relation features models.

Performed phenomenon-wide association studies (PheWa) and Phenotype Risk Scores (PheRS) have uncovered a plethora of diagnoses associated with LC. These studies associated seven phenotypes with the pre-COVID-19 period (e.g., irritable bowel syndrome, concussion, nausea/vomiting, and shortness of breath) and 69 acute-COVID-19 phenotypes (predominantly respiratory and circulatory phenotypes) significantly associated with LC. Using PheRS, a quarter of the COVID-19 positive cohort was identified with a 3.5-fold increased risk of LC compared to the bottom 50% of their distributions (121).

Sengupta et al. (122) proposed an interpretable DL approach based on Gradient-weighted Class Activation Mapping using N3C and RECOVER data to predict risk factors contributing to the development of LC. This model used a temporally ordered list of diagnostic codes six weeks post-COVID-19 infection for each patient, with an accuracy of 70.48%. Gupta et al. (123) proposed a stacking ensemble learning technique based on deep neural networks for early predicting cardiovascular disease risk in recovered SARS-CoV-2 patients with LC symptoms, achieving an accuracy of 93.23%.

The here reviewed studies highlight the versatility of ML methods to study LC, facilitating not only the implementation of predictive diagnostic tools but also encouraging the integration of clinical data with, social, demographic and other information, for the development of robust systems. Despite the versatility of ML techniques, there are still enormous challenges for their application in LC analysis, in particular the collection of meaningful data sets for the development of predictive systems.

3. Discussion

Mathematical models have helped to understand the dynamics of the spread of SARS-CoV-2 and helped to predict different scenarios during the COVID-19 pandemic, becoming one of the most relevant tools for developing public health policies. Correlating sanitary measures with virus variants and the effects on the reproduction rate enabled the assessment of government policies that will help to face new outbreaks of SARS-CoV-2 or future pandemics. The development of reliable mathematical models, statistical techniques for test correction, and methods of analysis of heterogeneous populations, together with the value of testing strategies and traceability of close contacts, has been remarkable achievements. Combining these systems with ML and AI methods increased the predictive power of the models and facilitated the simulation of scenarios.

Developing predictive systems for COVID-19 was one of the significant challenges assumed by thousands of scientists during the pandemic. The main achievements were developing models for clinical diagnostic systems, ML for drug and vaccine discovery, and forecasting models for epidemiological variables to support public health policies and monitoring systems. In turn, the development of predictive systems coupled with techniques such as protein language models and molecular techniques facilitated the study of variants at the genomic level. Such models helped to understand how mutations affected critical viral proteins, helping drug and vaccine designs.

The ongoing pandemic has introduced a complete set of challenges, and currently, a novel multisystem disease defined by the persistence or appearance of new symptoms after SARS-CoV-2 infection has emerged. This complex entity-denominated LC has yet to be fully elucidated, mainly because

it is characterized by a wide range of clinical manifestations, methodological limitations, and heterogeneous definitions that make clinical and computational analysis difficult. Despite rapidly emerging studies and growing evidence, current data needs to be improved. A primary task is to establish an approach to identify natural language data associated with potential LC patients. This task will likely require well-designed prospective studies, unified definitions of LC, an accurate distinction of SARS-CoV-2-related symptoms, and adequate follow-up times that include current patients, underrepresented groups, children, and minority populations. It is granted that ML strategies will play a critical role in the understanding of LC and other upcoming challenges of the ongoing SARS-CoV-2 pandemic.

Author contributions

LS, JG-P, and DM-O: conceptualization. DM-O, DA-S, and JA: methodology. DM-O and MN: validation. LS, JG-P, DM-O, JA, and DA-S: investigation. LS, DM-O, JG-P, and MN: writing, review, and editing. MN and RU-P: supervision, funding resources, and project administration. All authors contributed to the article and approved the submitted version.

Funding

The authors acknowledge funding by the MAG-2095 project, Ministry of Education, Chile. DM-O acknowledges ANID for the project SUBVENCIÓN A INSTALACIÓN EN LA ACADEMIA CONVOCATORIA AÑO 2022, Folio 85220004. MN acknowledges ANID for project ACT210085 and GORE Magallanes for project FIC-R 40036196-0.

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.

Publisher's note

References

1. Shankar S, Mohakuda SS, Kumar A, Nazneen P, Yadav AK, Chatterjee K, et al. Systematic review of predictive mathematical models of COVID-19 epidemic. *Med J Armed Forces India*. (2021) 77:S385–392. doi: 10.1016/j.mjafi.2021.05.005

2. Contreras S, Medina-Ortiz D, Conca C, Olivera-Nappa Á. A novel synthetic model of the glucose-insulin system for patient-wise inference of physiological parameters from small-size OGTT data. *Front Bioeng Biotechnol*. (2020) 8:195. doi: 10.3389/fbioe.2020.00195

3. Huang Y, Yang C, Xu XF, Xu W, Liu SW. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol Sin*. (2020) 41:1141–9. doi: 10.1038/s41401-020-0485-4

4. Magazine N, Zhang T, Wu Y, McGee MC, Veggiani G, Huang W. Mutations and evolution of the SARS-CoV-2 spike protein. *Viruses*. (2022) 14:640. doi: 10.3390/v14030640

5. Velavan TP, Meyer CG. The COVID-19 epidemic. *Trop Med Int Health*. (2020) 25:278. doi: 10.1111/tmi.13383

6. Baker MG, Wilson N. The covid-19 elimination debate needs correct data. *BMJ*. (2020) 371. doi: 10.1136/bmj.m3883

7. 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

8. Wu Z, Harrich D, Li Z, Hu D, Li D. The unique features of SARS-CoV-2 transmission: comparison with SARS-CoV, MERS-CoV and 2009 H1N1 pandemic influenza virus. *Rev Med Virol*. (2021) 31:e2171. doi: 10.1002/rmv.2171

9. Li Y, Campbell H, Kulkarni D, Harpur A, Nundy M, Wang X, et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. *Lancet Infect Dis*. (2021) 21:193–202. doi: 10.1016/S1473-3099(20)30785-4

10. Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Gavenčiak T, et al. Inferring the effectiveness of government interventions against COVID-19. *Science*. (2021) 371:eabd9338. doi: 10.1126/science.abd9338

11. Howard J, Huang A, Li Z, Tufekci Z, Zdimal V, van der Westhuizen HM, et al. An evidence review of face masks against COVID-19. *Proc Natl Acad Sci USA*. (2021) 118:e2014564118. doi: 10.1073/pnas.2014564118

12. Cobey S. Modeling infectious disease dynamics. *Science*. (2020) 368:713–714. doi: 10.1126/science.abb5659

13. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*. (2020) 584:257–61. doi: 10.1038/s41586-020-2405-7

14. Larremore DB, Wilder B, Lester E, Shehata S, Burke JM, Hay JA, et al. Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Sci Adv*. (2021) 7:eabd5393. doi: 10.1126/sciadv.abd5393

15. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell*. (2020) 182:812–27. doi: 10.1016/j.cell.2020.06.043

16. Fine PE. Herd immunity: history, theory, practice. *Epidemiol Rev*. (1993) 15:265–302. doi: 10.1093/oxfordjournals.epirev.a036121

17. To KKW, Hung IFN, Ip JD, Chu AWH, Chan WM, Tam AR, et al. Coronavirus disease 2019 (COVID-19) re-infection by a phylogenetically distinct severe acute respiratory syndrome coronavirus 2 strain confirmed by whole genome sequencing. *Clin Infect Dis*. (2020) 73:e2946–51. doi: 10.1093/cid/ciaa1275

18. Saad-Roy CM, Morris SE, Metcalf CJE, Mina MJ, Baker RE, Farrar J, et al. Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes. *Science*. (2021) 372:363–70. doi: 10.1126/science.abg8663

19. Tartof SY, Slezak JM, Fischer H, Hong V, Ackerson BK, Ranasinghe ON, et al. Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: a retrospective cohort study. *Lancet*. (2021) 398:1407–16. doi: 10.1016/S0140-6736(21)02183-8

20. Tang P, Hasan MR, Chemaitelly H, Yassine HM, Benslimane FM, Al Khatib HA, et al. BNT162b2 and mRNA-1273 COVID-19 vaccine effectiveness against the SARS-CoV-2 Delta variant in Qatar. *Nat Med*. (2021) 27:2136–43. doi: 10.1038/s41591-021-01583-4

21. Prunas O, Warren JL, Crawford FW, Gazit S, Patalon T, Weinberger DM, et al. Vaccination with BNT162b2 reduces transmission of SARS-CoV-2 to household contacts in Israel. *Science*. (2022) 375:1151–4. doi: 10.1126/science.abl4292

22. González-Puelma J, Aldridge J, Montes de Oca M, Pinto M, Uribe-Paredes R, Fernández-Goycoolea J, et al. Mutation in a SARS-CoV-2 haplotype from sub-Antarctic Chile reveals new insights into the SpikeâĂŹs dynamics. *Viruses*. (2021) 13:883. doi: 10.3390/v13050883

23. Acevedo ML, Gaete-Argel A, Alonso-Palomares L, de Oca MM, Bustamante A, Gaggero A, et al. Differential neutralizing antibody responses elicited by CoronaVac

and BNT162b2 against SARS-CoV-2 Lambda in Chile. *Nat Microbiol*. (2022) 7:524–9. doi: 10.1038/s41564-022-01092-1

24. Lopez Bernal J, Andrews N, Gower C, Gallagher E, Simmons R, Thelwall S, et al. Effectiveness of Covid-19 vaccines against the B. 1.617. 2 (Delta) variant. *N Engl J Med*. (2021) 385:585–94. doi: 10.1056/NEJMoa2108891

25. Kristiansen H, Gad HH, Eskildsen-Larsen S, Despres P, Hartmann R. The oligoadenylate synthetase family: an ancient protein family with multiple antiviral activities. *J Interferon Cytokine Res*. (2011) 31:41–7. doi: 10.1089/jir.2010.0107

26. Pulliam JR, van Schalkwyk C, Govender N, von Gottberg A, Cohen C, Groome MJ, et al. Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. *Science*. (2022) 376:eabn4947.

27. Mathieu E, Ritchie H, Ortiz-Ospina E, Roser M, Hasell J, Appel C, et al. A global database of COVID-19 vaccinations. *Nat Hum Behav*. (2021) 5:947–53. doi: 10.1038/s41562-021-01122-8

28. Hasell J. A cross-country database of COVID-19 testing. *Sci Data*. (2020) 7:345. doi: 10.1038/s41597-020-00688-8

29. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, et al. GISAIDåĂŹs role in pandemic response. *China CDC Weekly*. (2021) 3:1049. doi: 10.46234/ccdcw2021.255

30. Calafiore GC, Novara C, Possieri C. A modified SIR model for the COVID-19 contagion in Italy. In: *2020 59th IEEE Conference on Decision and Control (CDC)*. Jeju: IEEE (2020). p. 3889–94.

31. Annas S, Pratama MI, Rifandi M, Sanusi W, Side S. Stability analysis and numerical simulation of SEIR model for pandemic COVID-19 spread in Indonesia. *Chaos Solitons Fractals*. (2020) 139:110072. doi: 10.1016/j.chaos.2020.110072

32. Contreras S, Villavicencio HA, Medina-Ortiz D, Biron-Lattes JP, Olivera-Nappa Á. A multi-group SEIRA model for the spread of COVID-19 among heterogeneous populations. *Chaos Solitons Fractals*. (2020) 136:109925. doi: 10.1016/j.chaos.2020.109925

33. Contreras S, Biron-Lattes JP, Villavicencio HA, Medina-Ortiz D, Llanovarced-Kawles N, Olivera-Nappa Á. Statistically-based methodology for revealing real contagion trends and correcting delay-induced errors in the assessment of COVID-19 pandemic. *Chaos Solitons Fractals*. (2020) 139:110087. doi: 10.1016/j.chaos.2020.110087

34. Sanchez-Daza A, Medina-Ortiz D, Olivera-Nappa A, Contreras S. COVID-19 modeling under uncertainty: statistical data analysis for unveiling true spreading dynamics and guiding correct epidemiological management. In: *Modeling, Control and Drug Development for COVID-19 Outbreak Prevention*. Springer (2022). p. 245–82. Available online at: https://link.springer.com/chapter/10.1007/978-3-030-72834-2_9

35. Bauer S, Contreras S, Dehning J, Linden M, Iftekhar E, Mohr SB, et al. Relaxing restrictions at the pace of vaccination increases freedom and guards against further COVID-19 waves. *PLoS Comput Biol*. (2021) 17:e1009288. doi: 10.1371/journal.pcbi.1009288

36. Contreras S, Dehning J, Mohr SB, Bauer S, Spitzner FP, Priesemann V. Low case numbers enable long-term stable pandemic control without lockdowns. *Sci Adv*. (2021) 7:eabg2243. doi: 10.1126/sciadv.abg2243

37. Contreras S, Priesemann V. Risking further COVID-19 waves despite vaccination. *Lancet Infect Dis*. (2021) 21:745–6. doi: 10.1016/S1473-3099(21)00167-5

38. Oróstica KY, Contreras S, Mohr SB, Dehning J, Bauer S, Medina-Ortiz D, et al. Mutational signatures and transmissibility of SARS-CoV-2 Gamma and Lambda variants. *arXiv preprint* arXiv:210810018. (2021). doi: 10.48550/arXiv.2108.10018

39. Contreras S, Oróstica KY, Daza-Sanchez A, Wagner J, Dönges P, Medina-Ortiz D, et al. Model-based assessment of sampling protocols for infectious disease genomic surveillance. *Chaos Solitons Fractals*. (2023) 167:113093. doi: 10.1016/j.chaos.2022.113093

40. Oróstica KY, Contreras S, Sanchez-Daza A, Fernandez J, Priesemann V, Olivera-Nappa Á. New year, new SARS-CoV-2 variant: resolutions on genomic surveillance protocols to face Omicron. *Lancet Regional Health Am*. (2022) 7:100203. doi: 10.1016/j.lana.2022.100203

41. Hall VJ, Foulkes S, Saei A, Andrews N, Oguti B, Charlett A, et al. COVID-19 vaccine coverage in health-care workers in England and effectiveness of BNT162b2 mRNA vaccine against infection (SIREN): a prospective, multicentre, cohort study. *Lancet*. (2021) 397:1725–735. doi: 10.1016/S0140-6736(21)00790-X

42. Joshi G, Borah P, Thakur S, Sharma P, Mayank, Poduri R. Exploring the COVID-19 vaccine candidates against SARS-CoV-2 and its variants: where do we stand and where do we go? *Hum Vaccines Immunotherapeut*. (2021) 17:4714–40. doi: 10.1080/21645515.2021.1995283

43. Forni G, Mantovani A. COVID-19 vaccines: where we stand and challenges ahead. *Cell Death Diff*. (2021) 28:626–39. doi: 10.1038/s41418-020-00720-9

44. Dai L, Gao GF. Viral targets for vaccines against COVID-19. *Nat Rev Immunol*. (2021) 21:73–82. doi: 10.1038/s41577-020-00480-0

45. Rawat K, Kumari P, Saha L. COVID-19 vaccine: a recent update in pipeline vaccines, their design and development strategies. *Eur J Pharmacol*. (2021) 892:173751. doi: 10.1016/j.ejphar.2020.173751

46. Kyriakidis NC, López-Cortés A, González EV, Grimaldos AB, Prado EO. SARS-CoV-2 vaccines strategies: a comprehensive review of phase 3 candidates. *npj Vaccines*. (2021) 6:28. doi: 10.1038/s41541-021-00292-w

47. Baden L, El Sahly H, Essink B, et al. Učinkovitost in varnost cepiva mRNA-1273 SARS-CoV-2. *N Engl J Med*. (2021) 384:403–416. doi: 10.1056/NEJMoa2035389

48. Sa M, Bukhari IA, Akram J, Meo AS, Klonoff DC. COVID-19 vaccines: comparison of biological, pharmacological characteristics and adverse effects of Pfizer/BionTech and Moderna Vaccines. *Eur Rev Med Pharmacol Sci*. (2021) 25:1663–9. doi: 10.26355/eurrev_202102_24877

49. Klein NP, Lewis N, Goddard K, Fireman B, Zerbo O, Hanson KE, et al. Surveillance for adverse events after COVID-19 mRNA vaccination. *JAMA*. (2021) 326:1390–9. doi: 10.1001/jama.2021.15072

50. Shimabukuro T. Allergic reactions including anaphylaxis after receipt of the first dose of Pfizer-BioNTech COVID-19 vaccine–United States, December 14-23, 2020. *Am J Transpl*. (2021) 21:1332. doi: 10.1111/ajt.16516

51. Friedensohn L, Levin D, Fadlon-Derai M, Gershovitz L, Fink N, Glassberg E, et al. Myocarditis following a third BNT162b2 vaccination dose in military recruits in Israel. *JAMA*. (2022) 327:1611–2. doi: 10.1001/jama.2022.4425

52. See I, Su JR, Lale A, Woo EJ, Guh AY, Shimabukuro TT, et al. US case reports of cerebral venous sinus thrombosis with thrombocytopenia after Ad26. COV2. S vaccination, March 2 to April 21, 2021. *JAMA*. (2021) 325:2448–56. doi: 10.1001/jama.2021.7517

53. Hanson KE, Goddard K, Lewis N, Fireman B, Myers TR, Bakshi N, et al. Incidence of Guillain-Barré syndrome after COVID-19 vaccination in the vaccine safety datalink. *JAMA Network Open*. (2022) 5:e228879-e228879. doi: 10.1001/jamanetworkopen.2022.8879

54. Grome HN, Threlkeld M, Threlkeld S, Newman C, Martines RB, Reagan-Steiner S, et al. Fatal multisystem inflammatory syndrome in adult after SARS-CoV-2 natural infection and COVID-19 vaccination. *Emerg Infect Dis*. (2021) 27:2914. doi: 10.3201/eid2711.211612

55. Miao G, Chen Z, Cao H, Wu W, Chu X, Liu H, et al. From immunogen to COVID-19 vaccines: prospects for the post-pandemic era. *Biomed Pharmacother*. (2023) 2023:114208. doi: 10.1016/j.biopha.2022.114208

56. Castanares-Zapatero D, Chalon P, Kohn L, Dauvrin M, Detollenaere J, Maertens de Noordhout C, et al. Pathophysiology and mechanism of long COVID: a comprehensive review. *Ann Med*. (2022) 54:1473–87. doi: 10.1080/07853890.2022.2076901

57. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV, Group WCCDW, et al. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis*. (2021) 22:e102–7. doi: 10.1016/S1473-3099(21)00703-9

58. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and predictors of long COVID. *Nat Med*. (2021) 27:626–31. doi: 10.1038/s41591-021-01292-y

59. Boix V, Merino E. Post-COVID syndrome. The never ending challenge. *Med Clin*. (2022) 158:178. doi: 10.1016/j.medcle.2021.10.005

60. Yong SJ, Liu S. Proposed subtypes of post-COVID-19 syndrome (or long-COVID) and their respective potential therapies. *Rev Med Virol*. (2022) 32:e2315. doi: 10.1002/rmv.2315

61. Fernández-de Las-Pe nas C, Notarte KI, Peligro PJ, Velasco JV, Ocampo MJ, Henry BM, et al. Long-COVID symptoms in individuals infected with different SARS-CoV-2 variants of concern: a systematic review of the literature. *Viruses*. (2022) 14:2629. doi: 10.3390/v14122629

62. Stavem K, Ghanima W, Olsen MK, Gilboe HM, Einvik G. Persistent symptoms 1.5-6 months after COVID-19 in non-hospitalised subjects: a population-based cohort study. *Thorax*. (2021) 76:405–7. doi: 10.1136/thoraxjnl-2020-216377

63. Su Y, Yuan D, Chen DG, Ng RH, Wang K, Choi J, et al. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell*. (2022) 185:881–95. doi: 10.1016/j.cell.2022.01.014

64. Carfì A, Bernabei R, Landi F. Against COVID-19. Post-Acute Care Study Group: for the Gemelli Against CCOVID-19 Post-Acute Care Study Group. Persistent symptoms in patients after acute COVID-19. *JAMA*. (2020) 9:603. doi: 10.1001/jama.2020.12603

65. Dennis A, Wamil M, Alberts J, Oben J, Cuthbertson DJ, Wootton D, et al. Multiorgan impairment in low-risk individuals with post-COVID-19 syndrome: a prospective, community-based study. *BMJ Open*. (2021) 11:e048391. doi: 10.1136/bmjopen-2020-048391

66. Tanne JH. Covid-19: even mild infections can cause long term heart problems, large study finds. *Br Med J*. (2022) 2022:378. doi: 10.1136/bmj.o378

67. Qin W, Chen S, Zhang Y, Dong F, Zhang Z, Hu B, et al. Diffusion capacity abnormalities for carbon monoxide in patients with COVID-19 at 3-month follow-up. *Eur Respir J*. (2021) 58:2003677. doi: 10.1183/13993003.03677-2020

68. Sun B, Tang N, Peluso MJ, Iyer NS, Torres L, Donatelli JL, et al. Characterization and biomarker analyses of post-COVID-19 complications and neurological manifestations. *Cells*. (2021) 10:386. doi: 10.3390/cells10020386

69. Lamers MM, Beumer J, Van Der Vaart J, Knoops K, Puschhof J, Breugem TI, et al. SARS-CoV-2 productively infects human gut enterocytes. *Science*. (2020) 369:50–4. doi: 10.1126/science.abc1669

70. Bhattacharyya A, Seth A, Rai S. The effects of long COVID-19, its severity, and the need for immediate attention: analysis of clinical trials and Twitter data. *medRxiv*. (2022) 2022–09. doi: 10.1101/2022.09.13.22279833

71. Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, et al. The National COVID Cohort Collaborative (N3C): rationale, design, infrastructure, and deployment. *J Am Med Inform Assoc*. (2021) 28:427–43. doi: 10.1093/jamia/ocaa196

72. Alyasseri ZAA, Al-Betar MA, Doush IA, Awadallah MA, Abasi AK, Makhadmeh SN, et al. Review on COVID-19 diagnosis models based on machine learning and deep learning approaches. *Expert Syst*. (2022) 39:e12759. doi: 10.1111/exsy.12759

73. de Fátima Cobre A, Surek M, Stremel DP, Fachi MM, Borba HHL, Tonin FS, et al. Diagnosis and prognosis of COVID-19 employing analysis of patients' plasma and serum via LC-MS and machine learning. *Comput Biol Med*. (2022) 146:105659. doi: 10.1016/j.compbiomed.2022.105659

74. Kolozsvári LR, Bérczes T, Hajdu A, Gesztelyi R, Tiba A, Varga I, et al. Predicting the epidemic curve of the coronavirus (SARS-CoV-2) disease (COVID-19) using artificial intelligence: an application on the first and second waves. *Inform Med Unlocked*. (2021) 25:100691. doi: 10.1016/j.imu.2021.100691

75. Chandra R, Jain A, Singh Chauhan D. Deep learning via LSTM models for COVID-19 infection forecasting in India. *PLoS ONE*. (2022) 17:e0262708. doi: 10.1371/journal.pone.0262708

76. Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: a review. *Chaos Solitons Fractals*. (2020) 139:110059. doi: 10.1016/j.chaos.2020.110059

77. Zoabi Y, Deri-Rozov S, Shomron N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *npj Digit Med*. (2021) 4:1–5. doi: 10.1038/s41746-020-00372-6

78. Zhao L, Lediju Bell MA. A review of deep learning applications in lung ultrasound imaging of COVID-19 patients. *BME Front*. (2022) 2022:9780173. doi: 10.34133/2022/9780173

79. Dang T, Han J, Xia T, Spathis D, Bondareva E, Siegele-Brown C, et al. Exploring longitudinal cough, breath, and voice data for COVID-19 progression prediction via sequential deep learning: model development and validation. *J Med Internet Res*. (2022) 24:e37004. doi: 10.2196/37004

80. Jin W, Dong S, Dong C, Ye X. Hybrid ensemble model for differential diagnosis between COVID-19 and common viral pneumonia by chest X-ray radiograph. *Comput Biol Med*. (2021) 131:104252. doi: 10.1016/j.compbiomed.2021.104252

81. Soares E, Angelov P, Biaso S, Froes MH, Abe DK. SARS-CoV-2 CT-scan dataset: a large dataset of real patients CT scans for SARS-CoV-2 identification. *MedRxiv*. (2020). doi: 10.1101/2020.04.24.20078584

82. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. *TensorFlow*: a system for *Large-Scale* machine learning. In: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. (2016). p. 265–83. Available online at: https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

83. Alafif T, Tehame AM, Bajaba S, Barnawi A, Zia S. Machine and deep learning towards COVID-19 diagnosis and treatment: survey, challenges, and future directions. *Int J Environ Res Public Health*. (2021) 18:1117. doi: 10.3390/ijerph18031117

84. Lv H, Shi L, Berkenpas JW, Dao FY, Zulfiqar H, Ding H, et al. Application of artificial intelligence and machine learning for COVID-19 drug discovery and vaccine design. *Brief Bioinform*. (2021) 22:bbab320. doi: 10.1093/bib/bbab320

85. Law JN, Akers K, Tasnina N, Della Santina CM, Kshirsagar M, Klein-Seetharaman J, et al. Identifying human interactors of SARS-CoV-2 proteins and drug targets for COVID-19 using network-based label propagation. *arXiv preprint* arXiv:200601968. (2020).

86. Ray S, Lall S, Bandyopadhyay S. A deep integrated framework for predicting SARS-CoV2-human protein-protein interaction. *IEEE Trans Emerg Top Comput Intell*. (2022) 6:1463–72. doi: 10.1109/TETCI.2022.3182354

87. Du BX, Qin Y, Jiang YF, Xu Y, Yiu SM, Yu H, et al. Compound-protein interaction prediction by deep learning: databases, descriptors and models. *Drug Discov Today*. (2022) 27:1350–66. doi: 10.1016/j.drudis.2022.02.023

88. Mylonas SK, Axenopoulos A, Daras P. DeepSurf: a surface-based deep learning approach for the prediction of ligand binding sites on proteins. *Bioinformatics*. (2021) 37:1681–90. doi: 10.1093/bioinformatics/btab009

89. Liu Q, Wang PS, Zhu C, Gaines BB, Zhu T, Bi J, et al. OctSurf: efficient hierarchical voxel-based molecular surface representation for protein-ligand affinity prediction. *J Mol Graphics Model*. (2021) 105:107865. doi: 10.1016/j.jmgm.2021.107865

90. Jones D, Kim H, Zhang X, Zemla A, Stevenson G, Bennett WD, et al. Improved protein-ligand binding affinity prediction with structure-based deep fusion inference. *J Chem Inf Model*. (2021) 61:1583–92. doi: 10.1021/acs.jcim.0c01306

91. Feng Z, Chen M, Liang T, Shen M, Chen H, Xie XQ. Virus-CKB: an integrated bioinformatics platform and analysis resource for COVID-19 research. *Brief Bioinform*. (2021) 22:882–95. doi: 10.1093/bib/bbaa155

92. Zhavoronkov A, Aladinskiy V, Zhebrak A, Zagribelnyy B, Terentiev V, Bezrukov D, et al. Potential COVID-2019 3C-like protease inhibitors designed using generative deep learning approaches. *ChemRxiv. Preprint*. (2020) 11:102. doi: 10.26434/chemrxiv.11829102

93. Tang B, He F, Liu D, He F, Wu T, Fang M, et al. AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2. *Biomolecules*. (2022) 12:746. doi: 10.3390/biom12060746

94. Ton AT, Gentile F, Hsing M, Ban F, Cherkasov A. Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. *Mol Inform*. (2020) 39:2000028. doi: 10.1002/minf.202000028

95. Srinivasan S, Batra R, Chan H, Kamath G, Cherukara MJ, Sankaranarayanan SK. Artificial intelligence-guided De novo molecular design targeting COVID-19. *ACS Omega*. (2021) 6:12557–66. doi: 10.1021/acsomega.1c00477

96. Wang Z, Liu M, Luo Y, Xu Z, Xie Y, Wang L, et al. Advanced graph and sequence neural networks for molecular property prediction and drug discovery. *Bioinformatics*. (2022) 38:2579–86. doi: 10.1093/bioinformatics/btac112

97. Medina-Ortiz D, Contreras S, Amado-Hinojosa J, Torres-Almonacid J, Asenjo JA, Navarrete M, et al. Generalized property-based encoders and digital signal processing facilitate predictive tasks in protein engineering. *Front Mol Biosci*. (2022) 9:898627. doi: 10.3389/fmolb.2022.898627

98. Ferruz N, Höcker B. Controllable protein design with language models. *Nat Mach Intell*. (2022) 4:521–32. doi: 10.1038/s42256-022-00499-z

99. Wigh DS, Goodman JM, Lapkin AA. A review of molecular representation in the age of machine learning. *Wiley Interdisc Rev Comput Mol Sci*. (2022) 12:e1603. doi: 10.1002/wcms.1603

100. Verkuil R, Kabeli O, Du Y, Wicky BI, Milles LF, Dauparas J, et al. Language models generalize beyond natural proteins. *bioRxiv*. (2022) doi: 10.1101/2022.12.21.521521

101. Minh D, Wang HX, Li YF, Nguyen TN. Explainable artificial intelligence: a comprehensive review. *Artif Intell Rev*. (2022) 55:3503–68. doi: 10.1007/s10462-021-10088-y

102. Chen CC, Chang CK, Chiu CC, Yang TY, Hao WR, Lin CH, et al. Machine learning analyses revealed distinct arterial pulse variability according to side effects of Pfizer-BioNTech COVID-19 vaccine (BNT162b2). *J Clin Med*. (2022) 11:6119. doi: 10.3390/jcm11206119

103. Abbaspour S, Robbins GK, Blumenthal KG, Hashimoto D, Hopcia K, Mukerji SS, et al. Identifying modifiable predictors of COVID-19 vaccine side effects: a machine learning approach. *Vaccines*. (2022) 10:1747. doi: 10.3390/vaccines10101747

104. Flora J, Khan W, Jin J, Jin D, Hussain A, Dajani K, et al. Usefulness of vaccine adverse event reporting system for machine-learning based vaccine research: A Case study for COVID-19 vaccines. *Int J Mol Sci*. (2022) 23:8235. doi: 10.3390/ijms23158235

105. Gupta M, Jain R, Taneja S, Chaudhary G, Khari M, Verdú E. Real-time measurement of the uncertain epidemiological appearances of COVID-19 infections. *Appl Soft Comput*. (2021) 101:107039. doi: 10.1016/j.asoc.2020.107039

106. Mansour RF, Escorcia-Gutierrez J, Gamarra M, Gupta D, Castillo O, Kumar S. Unsupervised deep learning based variational autoencoder model for COVID-19 diagnosis and classification. *Pattern Recogn Lett*. (2021) 151:267–74. doi: 10.1016/j.patrec.2021.08.018

107. Wang D, Zuo F, Gao J, He Y, Bian Z, Bernardes SD, et al. Agent-based simulation model and deep learning techniques to evaluate and predict transportation trends around COVID-19. *arXiv preprint* arXiv:201009648. (2020). doi: 10.48550/arXiv.2010.09648

108. Kompella V, Capobianco R, Jong S, Browne J, Fox S, Meyers L, et al. Reinforcement learning for optimization of COVID-19 mitigation policies. *arXiv preprint* arXiv:201010560. (2020). doi: 10.48550/arXiv.2010.10560

109. Medina-Ortiz D, Contreras S, Barrera-Saavedra Y, Cabas-Mora G, Olivera-Nappa Á. Country-wise forecast model for the effective reproduction number R t of coronavirus disease. *Front Phys*. (2020) 8:304. doi: 10.3389/fphy.2020.00304

110. Contreras S, Villavicencio HA, Medina-Ortiz D, Saavedra CP, Olivera-Nappa Á. Real-time estimation of R t for supporting public-health policies against COVID-19. *Front Public Health*. (2020). 8:556689. doi: 10.3389/fpubh.2020.556689

111. Polyzos S, Samitas A, Spyridou AE. Tourism demand and the COVID-19 pandemic: an LSTM approach. *Tour Recreat Res*. (2021) 46:175–87. doi: 10.1080/02508281.2020.1777053

112. Hills S, Eraso Y. Factors associated with non-adherence to social distancing rules during the COVID-19 pandemic: a logistic regression analysis. *BMC Public Health*. (2021) 21:1–25. doi: 10.1186/s12889-021-10379-7

113. Saba AI, Elsheikh AH. Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks. *Process Safety Environ Protect*. (2020) 141:1–8. doi: 10.1016/j.psep.2020.05.029

114. Singh RK, Rani M, Bhagavathula AS, Sah R, Rodriguez-Morales AJ, Kalita H, et al. Prediction of the COVID-19 pandemic for the top 15 affected countries: Advanced autoregressive integrated moving average (ARIMA) model. *JMIR Public Health Surveill*. (2020) 6:e19115. doi: 10.2196/19115

115. Ketu S, Mishra PK. Enhanced Gaussian process regression-based forecasting model for COVID-19 outbreak and significance of IoT for its detection. *Appl Intell*. (2021) 51:1492–512. doi: 10.1007/s10489-020-01889-9

116. Castillo Ossa LF, Chamoso P, Arango-López J, Pinto-Santos F, Isaza GA, Santa-Cruz-González C, et al. A hybrid model for COVID-19 monitoring and prediction. *Electronics*. (2021) 10:799. doi: 10.3390/electronics10070799

117. Cheng FY, Joshi H, Tandon P, Freeman R, Reich DL, Mazumdar M, et al. Using machine learning to predict ICU transfer in hospitalized COVID-19 patients. *J Clin Med*. (2020) 9:1668. doi: 10.3390/jcm9061668

118. Alamrouni A, Aslanova F, Mati S, Maccido HS, Jibril AA, Usman A, et al. Multi-regional modeling of cumulative COVID-19 cases integrated with environmental forest knowledge estimation: a deep learning ensemble approach. *Int J Environ Res Public Health*. (2022) 19:738. doi: 10.3390/ijerph19020738

119. Pfaff ER, Girvin AT, Bennett TD, Bhatia A, Brooks IM, Deer RR, et al. Identifying who has long COVID in the USA: a machine learning approach using N3C data. *Lancet Digit Health*. (2022) 4:e532–41. doi: 10.1016/S2589-7500(22)00048-6

120. Binka M, Klaver B, Cua G, Wong AW, Fibke C, Velásquez García HA, et al. An elastic net regression model for identifying long COVID patients using health administrative data: a population-based study. In: *Open Forum Infectious Diseases. vol. 9*. Oxford: Oxford University Press US (2022). p. ofac640.

121. Fritsche LG, Jin W, Admon AJ, Mukherjee B. Characterizing and predicting post-acute sequelae of SARS CoV-2 infection (PASC) in a large academic medical center in the US. *medRxiv*. (2022) doi: 10.1101/2022.10.21.22281356

122. Sengupta S, Loomba J, Sharma S, Brown DE, Thorpe L, Haendel MA, et al. Analyzing historical diagnosis code data from NIH N3C and RECOVER Programs using deep learning to determine risk factors for Long COVID. *arXiv preprint* arXiv:221002490. (2022) doi: 10.1109/BIBM55620.2022.9994851

123. Gupta A, Jain V, Singh A. Stacking ensemble-based intelligent machine learning model for predicting post-COVID-19 complications. *N Generat Comput*. (2022) 40:987–1007. doi: 10.1007/s00354-021-00144-0

frontiers | Frontiers in Public Health

TYPE Review
PUBLISHED 11 May 2023
DOI 10.3389/fpubh.2023.1126461

OPEN ACCESS

EDITED BY
Pierpaolo Ferrante,
National Institute for Insurance against Accidents at Work (INAIL), Italy

REVIEWED BY
Jorge P. Rodríguez,
Instituto Mediterráneo de Estudios Avanzados (IMEDEA), CSIC-UIB, Spain
Flávio Codeço Coelho,
Fundação Getúlio Vargas, Brazil

*CORRESPONDENCE
Igor T. Peres
✉ igor.peres@puc-rio.br

†These authors have contributed equally to this work

RECEIVED 17 December 2022
ACCEPTED 06 April 2023
PUBLISHED 11 May 2023

CITATION
Hastenreiter Filho HN, Peres IT, Maddalena LG, Baião FA, Ranzani OT, Hamacher S, Maçaira PM and Bozza FA (2023) What we talk about when we talk about COVID-19 vaccination campaign impact: a narrative review.
Front. Public Health 11:1126461.
doi: 10.3389/fpubh.2023.1126461

What we talk about when we talk about COVID-19 vaccination campaign impact: a narrative review

Horácio N. Hastenreiter Filho[1,2†], Igor T. Peres[1*†], Lucas G. Maddalena[1], Fernanda A. Baião[1], Otavio T. Ranzani[3,4], Silvio Hamacher[1], Paula M. Maçaira[1] and Fernando A. Bozza[5,6]

[1]Department of Industrial Engineering, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil, [2]School of Management, Federal University of Bahia, Salvador, Brazil, [3]Barcelona Institute for Global Health, Barcelona, Spain, [4]Pulmonary Division, Heart Institute, Faculty of Medicine, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil, [5]National Institute of Infectious Disease Evandro Chagas, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil, [6]D'Or Institute for Research and Education, Rio de Janeiro, Brazil

Background: The lack of precise definitions and terminological consensus about the impact studies of COVID-19 vaccination leads to confusing statements from the scientific community about what a vaccination impact study is.

Objective: The present work presents a narrative review, describing and discussing COVID-19 vaccination impact studies, mapping their relevant characteristics, such as study design, approaches and outcome variables, while analyzing their similarities, distinctions, and main insights.

Methods: The articles screening, regarding title, abstract, and full-text reading, included papers addressing perspectives about the impact of vaccines on population outcomes. The screening process included articles published before June 10, 2022, based on the initial papers' relevance to this study's research topics. The main inclusion criteria were data analyses and study designs based on statistical modelling or comparison of pre- and post-vaccination population.

Results: The review included 18 studies evaluating the vaccine impact in a total of 48 countries, including 32 high-income countries (United States, Israel, and 30 Western European countries) and 16 low- and middle-income countries (Brazil, Colombia, and 14 Eastern European countries). We summarize the main characteristics of the vaccination impact studies analyzed in this narrative review.

Conclusion: Although all studies claim to address the impact of a vaccination program, they differ significantly in their objectives since they adopt different definitions of impact, methodologies, and outcome variables. These and other differences are related to distinct data sources, designs, analysis methods, models, and approaches.

KEYWORDS

vaccine impact, narrative review, COVID-19, vaccination campaign, populational studies

1. Introduction

Since 2020, epidemiological studies related to the effects of vaccination against COVID-19 have been gaining prominence in leading international journals, reaching more than 700 studies in the Scopus database in June 2022. These papers apply distinct study designs and address different measures of vaccine performance. Clinical trials first stood out in the search to present the efficacy of the vaccines during their phase-3 periods before licensing for application in the general population.

With the beginning of the vaccination roll-outs worldwide, several researchers were dedicated to evaluating the vaccine's effects on individuals or populations. Vaccine efficacy is determined by randomized controlled trials, and vaccine effectiveness is estimated from post-introduction observational studies. While effectiveness and efficacy of vaccination measure the direct effect of a vaccine on the vaccinated individuals and aim to describe an individual's risk reduction after vaccination, studies on vaccine impact address the outcome of a vaccination program in a community. These studies are typically ecological or modeling analyses that compare disease outcomes from pre- and post-vaccine introduction. The reductions in disease outcomes are estimated through the direct effects of vaccination in vaccinated participants and indirect effects due to reduced transmission within a community (1).

Most vaccination efficacy studies assess an individual's risk reduction after being vaccinated compared with those unvaccinated, thus inevitably addressing vaccine effectiveness (2–4). Vaccination impact studies are typically more feasible since individualized data are not always available in many scenarios. Only aggregated or deidentified data about the vaccination progress is often publicly available to infer how the vaccination roll-out impacts the population. While vaccine effectiveness studies are more consistent in study design and estimates (5), the existing impact studies differ significantly in many perspectives, including different study designs, estimated community outcomes, confounder variables, data sources, methods and models.

Moreover, there are literature works that address the impact of COVID-19 vaccination but should be characterized as vaccine effectiveness studies instead. For example, while the title of the work by Pritchard et al. (6) mentions vaccine impact, it presented the reduction of individual infections in vaccinated people. In the same way, the main results shown in Tande et al. (7) refer to the relative risks between vaccinated and non-vaccinated individuals, and Moghadas et al. (8) presented a theoretical simulation addressing individual outcomes.

The lack of precise definitions and terminological consensus leads to confusing statements from the scientific community about what a vaccination impact study is. In addition, there is a myriad of possible study designs in the literature that address the impact of vaccination programs on distinct populations. Difficulties in comparing study results reduce the understanding of the potential impact of a COVID-19 vaccination program on a specific population.

The present work presents a narrative review describing and discussing COVID-19 vaccine impact studies, mapping their relevant characteristics, such as study design, approach and outcome variables, while analyzing their similarities, distinctions, and main insights. Our search approach not only aims to make explicit the real distinction between vaccination impact studies and vaccination efficacy and effectiveness studies, but also presents a range of possibilities of scope and methods, among other variables. The methodology applied here, followed by other recent publications (9–13), can be used to explain the impact of a vaccination roll-out in a community, guiding and equipping other researchers interested in the subject.

2. Materials and methods

We performed the electronic search using the PubMed and Google Scholar databases. The search included literature published before November 30, 2021, using the keywords "covid-19," "SARS-CoV-2," "vaccine*" and "impact*." The articles' screening included studies addressing the impact of COVID-19 vaccines on population outcomes. The study selection was conducted by: (i) formulating the eligibility criteria; (ii) reading the abstract and selecting for full-text reading; (iii) reading the full-texts and selecting for study inclusion; and (iv) conducting a snowballing process including other studies by forward and backward search (14–17).

We considered the following eligibility criteria for study inclusion: articles covering the vaccine impact research topic, with a design of statistical modeling and/or comparison between pre- and post-vaccination population, and written in English. We excluded articles not covering the topic of COVID-19 vaccination impact, lacking a detailed description, using unclear methods, addressing vaccine effectiveness rather than impact, or related to vaccine acceptance and health impacts. Among those who focused on the vaccine's effect on preventing COVID-19 cases, most presented the impact from the individual perspective of those vaccinated (effectiveness) and not the effect of the vaccination process on the entire population (vaccine impact). We further conducted a snowballing process from the first set of articles, including the literature before June 10, 2022. The snowballing search is a backward and forward screening, looking at the reference list of the included articles (backward) and the papers citing the included studies (forward) (18). We extracted the following research characteristics to be discussed: author, year, country, period of analysis, data source, design, outcomes and methods/models/approaches applied.

3. Results

The database search found 741 articles, which were screened following the eligibility criteria. From the title and abstract reading, we excluded 704 papers presenting unclear methods lacking a detailed description or not covering the topic of COVID-19 vaccination impact, and 25 studies analyzing vaccine effectiveness rather than vaccine impact. This first study selection process yielded a set of 12 articles. From the full-text reading, we excluded four papers that did not perform statistical modelling or comparisons between pre- and post-vaccination, leaving a total of eight articles. From the selected papers, we further conducted a snowballing search, finding 269 additional articles citing or being cited by those that composed the initial set. These additional articles were screened with the criteria described before, and 10 new articles were selected, totaling 18 articles. Figure 1 illustrates the flowchart of the whole screening process. As the objective was not to conduct a systematic review, articles with similar approaches covering the same countries

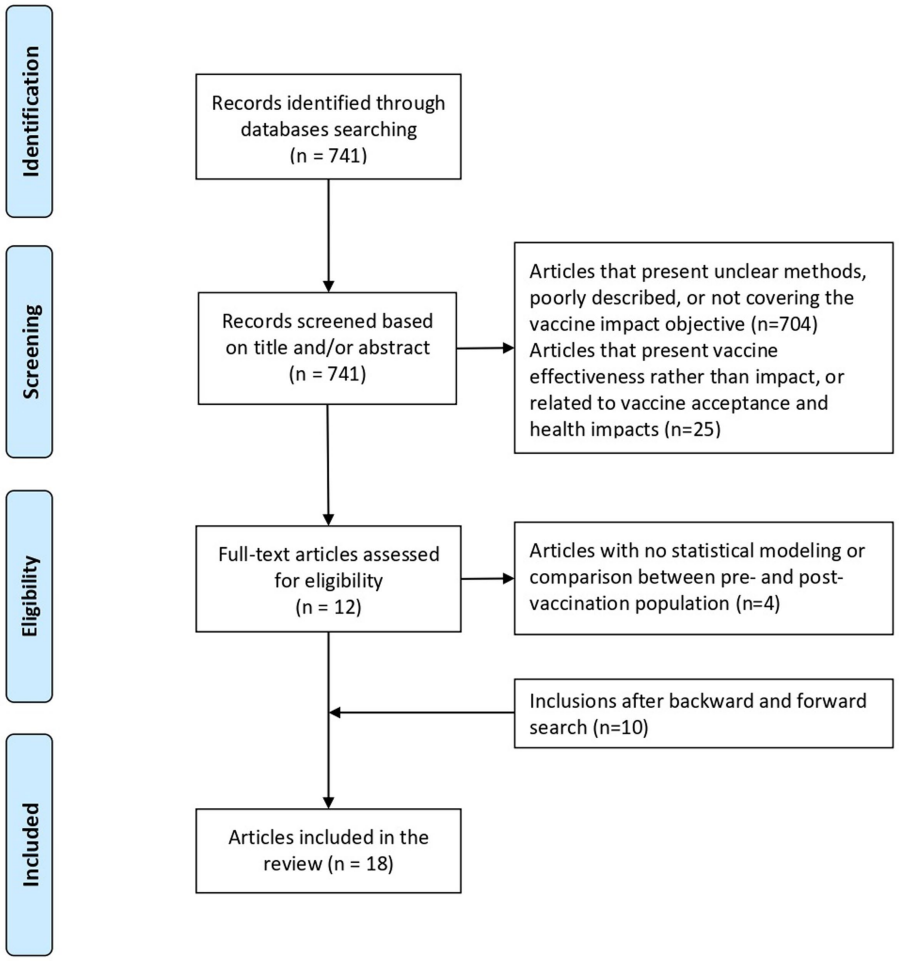


FIGURE 1
Flowchart of the screening process.

whose titles did not cover the vaccine impact theme were not analyzed.

The selected papers presented a diversity of methods, models, and approaches to address the impact of vaccines. Table 1 presents the characteristics of each article in terms of country, period, data source, design, outcomes and methods/models/approaches. The periods of reference for most studies involve 5 months or more. In some cases, the periods of reference for the used data start before the beginning of the vaccination to assess the evolution of cases before and after vaccination. There are studies from Latin America (Brazil and Colombia), North America (United States), Europe (England, Italy, Portugal, and other European countries) and Asia (Israel).

3.1. Summary of studies

Although the selected studies address the impact of a vaccination program, they differ significantly in their objective since they adopt different definitions of impact and methodologies. For instance, some studies have compared COVID-19 outcomes during different periods of the pandemic roll-out (specifically, the pre- and post-vaccination). In contrast, other studies perform a counterfactual analysis to calculate the vaccination program's impact on a population, estimating what could have been the COVID-19 outcome if either no vaccination program existed, or vaccination uptake had lower levels on the studied population.

Cot et al. (19) built an epidemic Renormalization Group (eRG) framework to reproduce and predict the diffusion of the pandemic in the U.S., taking human mobility across the U.S. and the influence of social distancing into account. Human mobility is monitored using open-source flight data among U.S. states. The eRG framework provides a single first-order differential equation that describes the time-evolution of the cumulative number of infected cases in an isolated region. Meslé et al. (20) estimated the number of deaths directly averted in the population of older adults (60 years and older) due to COVID-19 vaccination in the WHO European Region from December 2020 to November 2021. The authors simulated COVID-19 outcomes in a scenario without vaccination. The simulation parameters were based on information from previous studies of COVID-19 vaccine effectiveness in preventing deaths, thus calculating the number of directly averted deaths for each country. The analysis also applied an adapted formula used by Machado et al. (21) to measure the influenza vaccine program impact, which calculates the number of deaths averted with one dose and with full vaccination

TABLE 1 Main characteristics of included studies.

Author, year	Country	Period	Data sources	Study design	Outcomes	Method/models/ approaches applied
Cot et al., 2021	US	Dec 2020–Mar 2021	The OpenSky COVID-19 Flight Dataset and Opendatasoft	Epidemiological statistical modeling	COVID-19 case incidence	Epidemic Renormalization Group (eRG)
Mesle et al., 2021	World Health Organization (WHO) European Region	Dec 2020–Nov 2021	The European Surveillance System (TESSy)	Comparison of pre- and post-vaccination population	Averted COVID-19 deaths	Estimation of averted events
McNamara et al., 2022	US	Nov 2020–Apr 2021	Centers for Disease Control and Prevention (CDC)	Comparison of pre- and post-vaccination population	COVID-19 deaths	Difference-in-differences framework
Victora et al., 2021	Brazil	Jan–May2021	Brazilian Ministry of Health System	Comparison of pre- and post-vaccination population	COVD-19 mortality rate	Calculus of COVID-19 age-specific mortality rates
Rossman et al., 2021	Israel	Aug 2020–Feb 2021	Israeli Ministry of Health	Comparison of pre- and post-vaccination population	COVID-19 cases, hospitalizations, and severe hospitalizations	Temporal changes in weekly numbers of several clinical measures
Galvani et al., 2021	US	Oct 2020–Jun 2021	Centers for Disease Control and Prevention (CDC)	Epidemiological statistical modeling	Averted COVID-19 hospitalizations and deaths	Estimation of averted events
Andrews et al., 2021	England	Dec 2020–Mar 2021	Centers for Disease Control and Prevention (CDC)	Epidemiological statistical modeling	Averted COVID-19 hospitalizations and deaths	Estimation of averted events
Machado et al., 2022	Portugal	Dec 2020–Jul 2021	Portugal Health General Office	Epidemiological statistical modeling	Dynamics of confirmed cases and transmissibility index value (Rt)	SEIR model
Haas et al., 2022	Israel	Dec 2020–Apr 2021	National surveillance data from the Israeli Ministry of Health	Comparison of pre- and post-vaccination population	Averted SARS-CoV-2 infections and COVID-19- hospitalizations, severe hospitalizations, and deaths.	Estimation of averted events
Milman et al., 2021	Israel	Dec 2020–Mar 2021	Maccabi Healthcare Services database	Comparison of pre- and post-vaccination population	Relative changes in positive test fraction according to changes in the fraction vaccinated	Correlation analysis
Miłobedzki, 2022	European Union countries	Jan–Jul 2021	Our World in Data	Epidemiological statistical modeling	COVID-19 mortality	Estimation of confirmed new deaths based on infections and vaccinations
Liu et al., 2021	13 middle-income countries (MICs) of Europe.	Mar–Nov 2021	WHO Strategic Advisory Group of Experts on Immunization (SAGE) dataset	Epidemiological statistical modeling	COVID-19 mortality	Transmission Dynamic Model (adapted CovidM)
Caetano et al., 2021	Portugal	Jan–Sep 2021	ACSS/SPMS hospitalization registry	Comparison of pre- and post-vaccination population	COVID-19 averted deaths; Vaccine Effectiveness	SEIR model
Rojas-Botero et al., 2022	Colombia	Mar–Dec 2021	PAIWEB information system of the Ministry of Health and Social Protection	Comparison of pre- and post-vaccination population	COVID-19 averted deaths	Estimation of averted events; Estimation of Vaccine Effectiveness

(Continued)

10.3389/fpubh.2023.1126461

TABLE 1 (Continued)

Author, year	Country	Period	Data sources	Study design	Outcomes	Method/models/ approaches applied
Sacco et al., 2021	Italy	Jan–Sep 2021	Case- based national COVID-19 integrated surveillance system	Comparison of pre- and post-vaccination population	COVID-19 averted cases, hospitalizations, ICU admissions and deaths; Vaccine Effectiveness	Estimation of averted events; Estimation of Vaccine Effectiveness
Mattiuzzi et al., 2021	Europe (different countries)	Dec 2020–Nov 2021	Data of Meslé et al. (20)	Epidemiological statistical modeling	Association between the percentage of averted deaths of older people and percentage of vaccine uptake in each corresponding European country	Spearman's correlation and multiple linear regression
Shoukat et al., 2021	US	Dec 2020–Jul 2021	Centers for Disease Control and Prevention (CDC)	Epidemiological statistical modeling	Averted COVID-19 hospitalizations and deaths	Age-stratified agent-based model of COVID-19
Suthar et al., 2021	US	Dec 2020–Dec 2021	Centers for Disease Control and Prevention (CDC)	Epidemiological statistical modeling	COVID-19 case incidence and mortality	Generalized linear mixed models

through two different equations. The equations associate death numbers with vaccine effectiveness and vaccination uptake.

McNamara et al. (22) estimated the national-level impact of the initial phases of the COVID-19 vaccination program in the US. The authors compared relative changes in four different outcomes considering pre- and post-vaccination periods for the whole population and age groups. The authors applied a difference-in-differences framework to evaluate whether outcomes declined rapidly after vaccination roll-out in age groups with earlier vaccine eligibility. McNamara et al. (22) is mentioned by Ortiz and Neuzil (1) as an example of a COVID-19 vaccination program impact study. Victora et al. (23) investigated whether vaccination impacts the mortality of older individuals in a context of SARS-CoV-2 gamma variant (P.1 lineage) dominance in Brazil. The study analyzed the changes in COVID-19 proportionate mortality and mortality rate ratio in different age groups during the increase of vaccination coverage. First, they obtained proportionate mortality for older individuals (i.e., the ratio between the number of COVID-19 deaths at ages 70–79 and 80+ years and total number of COVID-19 deaths). Second, they calculated COVID-19 age-specific mortality rates by dividing the numbers of weekly deaths by the estimated population by age group. Mortality rates at ages 70–79 and 80+ years were then divided by rates for the age range 0–9 years in the same week, resulting in mortality rate ratios.

Rossman et al. (24) analyzed the temporal dynamics of new COVID-19 cases and hospitalizations after the vaccination campaign to distinguish the possible impact of vaccination from other factors, including a third lockdown implemented in Israel in January 2021. The authors performed several comparisons: individuals aged 60 years and older were prioritized to receive the vaccine first versus younger age groups; the January 2021 lockdown versus the September 2020 lockdown; and early vaccinated versus late-vaccinated cities. Galvani et al. (25) estimated the impact of the US COVID-19 vaccination campaign in controlling the virus's transmission and deaths. The authors compared COVID-19 outcomes on the current scenario with two counterfactuals: 50% of vaccination coverage and without a vaccination campaign. They estimated the averted number of COVID-19 deaths and hospitalizations, and calculated the adjusted odds ratios for vaccination impact, stratified by vaccine platform and previous SARS-COV-2 infection. To evaluate the vaccination program impact in the US, the researchers expanded their COVID-19 age-stratified agent-based model to include transmission dynamics of the different variants. They also used the population demographics, the contact network accounting for pandemic mobility patterns, and age-specific risks of severe health outcomes due to COVID-19 as model parameters.

Andrews et al. (26) estimated the number of deaths prevented by vaccination in England between the start of the vaccination program and the end of March 2021. Assessments are made to compare the COVID-19 mortality in the current scenario with an estimated counterfactual scenario without a vaccination program. Machado et al. (27) analyzed the impact of vaccination on the control of the pandemic. They investigated the relationship between vaccine coverage and non-pharmacological interventions (NPIs), developing different scenarios for the fade-out of NPIs as vaccine coverage increases in the population. The analysis is based on developing a standard mathematical model for assessing the population-level impact of a COVID-19 vaccine in a community. A SEIR model is created by splitting the total human population into mutually exclusive

compartments: unvaccinated susceptible vaccinated, susceptible, early exposed, pre-symptomatic infected, symptomatically infected, asymptomatically-infected, hospitalized and recovered.

Haas et al. (28) analyzed the number of averted COVID-19 infections, hospitalizations, and deaths in Israel due to the nationwide vaccination campaign using the Pfizer-BioNTech BNT162b2 mRNA COVID-19 vaccine. The authors estimated the direct effects of the immunization program for all susceptible individuals who were at least with one dose of COVID-19 vaccine compared to unvaccinated individuals. Moreover, Milman et al. (29) analyzed the community-level evidence for SARS-CoV-2 vaccine protection of unvaccinated individuals using a correlation analysis to test results collected during the rapid vaccine rollout in a large population from 177 Israeli communities. To control for the spatiotemporally dynamic nature of the epidemic, they focused on relative changes in the proportion of positive tests within each community between fixed time intervals.

Miłobedzki et al. (30) estimated the number of confirmed new deaths based on infections and vaccinations for the European Union countries. They computed the long-run marginal death effect concerning confirmed infections and compared it with respect to confirmed vaccinations. The authors also calculated the minimal weekly number of new vaccinations per million population in a European country to keep the number of new deaths per million population at a certain level. Liu et al. (31) applied a dynamic transmission model to analyze possible dosing interval strategies for two-dose COVID-19 vaccination in thirteen European middle-income countries and compared their impacts in terms of mortality. A vaccine with similar characteristics to AstraZeneca (AZD1222) was used in the base scenario. The authors also included sensitivity analyses considering different values for vaccine efficacy.

Caetano et al. (32) estimated the COVID-19 averted deaths in Portugal using a SEIR model to measure the impact of vaccination strategy. The authors adapted an age-structured SEIR deterministic model and used hospitalization data for the model calibration to measure the impact of the COVID-19 Portuguese vaccination strategy on the effective reproduction number. They also explored three scenarios for vaccine effectiveness waning: the no-immunity-loss, 1-year and 3-year immunity duration scenarios. Rojas-Botero et al. (33) estimated the number of directly averted deaths due to COVID-19 vaccination among older adults in Colombia. The authors calculated the full vaccination coverage of older adults, for each epidemiological week and age group, from March to December 2021. A sensitivity analysis considered variations in vaccine effectiveness by age group. Sacco et al. (34) estimated the number of averted COVID-19 cases, hospitalizations, intensive care unit admissions, and deaths by COVID-19 vaccination in Italy. The authors applied a method widely used in the study of vaccination impact during the influenza season (21, 35).

Mattiuzzi et al. (36) measure the association between the percentage of averted deaths of older people and the percentage of vaccine uptake in each corresponding European country. The authors used data on vaccine uptake and efficacy to perform univariate (Spearman's correlation) and multivariate (multiple linear regression analysis) correlations to determine the association of the percentage of averted deaths with vaccine uptake and the type of vaccine administered. Shoukat et al. (37) applied an age-stratified agent-based model of COVID-19 in US data to estimate the averted COVID-19 hospitalizations and deaths due to the vaccination roll-out. The model was calibrated using reported incidence in New York City (NYC), considering the relative transmissibility of each variant and vaccination coverage. The authors simulated the COVID-19 outbreak in NYC under the counterfactual scenario of no-vaccination and compared the resulting disease burden using the number of cases, hospitalizations, and deaths reported under the actual vaccination status. Also in US, Suthar et al. (38) used generalized linear mixed models assuming a negative binomial outcome distribution to analyze the impact of vaccines in reducing COVID-19 incidence and mortality. The authors also included a first-order autoregressive correlation structure to account for multiple observations per municipality and to identify potential autocorrelation.

The set of studies herein described sought to establish causal relationships between the vaccination process and different outcomes related to COVID-19. However, in Cot et al. (19) and Rossman et al. (24), there is the intermediation of confounders variables such as mobility and non-pharmacological interventions (NPI). Studies based exclusively on simulations, such as the one from Iboi et al. (39), were not included. Although all studies aimed to estimate the impact of the vaccination roll-out in a population-level, they used different analysis methods, which implies diverse models and tools, to achieve their established objectives. For instance, while Meslé et al. (20) estimated vaccination campaigns' impact by calculating the number of averted deaths, the study by McNamara et al. (22) estimated by comparing pre-vaccination COVID-19 outcomes with post-vaccination outcomes. In this sense, the study by Fang et al. (40) used the association between the vaccination coverage and the incidences and deaths caused by COVID-19 to calculate the impact of each percentage increase in population vaccination rates in the reduction of county-wide COVID-19 incidence and mortality. Often, the analysis method explained how the explanatory and outcome variables were associated. The differences among the analysed studies regarding their objectives lead to significant contrasts in the analysis methods, tools, and variables considered.

3.2. Data sources

The most important data for the studies are those related to the COVID-19 vaccination campaign, the confirmed cases and their outcomes. Usually, the National Ministry of Health and the Centers for Diseases Control are the main sources of these data. Nonetheless, depending on the approaches applied, other sources (secondary data) are also considered, as in Mattiuzzi et al. (36), which used the data produced by Meslé et al. (20).

3.3. Study design

By analyzing the populational level of the data used in the studies and their observational nature, we can say that all the studies follow an ecological study design, according to Levin et al. (41). More specifically, and according to Hanquet et al. (42), the impact of a vaccination program is estimated by comparing the population with access to a vaccination program with a reference population without the program, and vaccination program impact studies may follow mainly three different designs, which are specific subtypes of an ecological study:

- *Comparison of pre- and post-vaccination population.* According to this design, the two populations being assessed are separated by time, and the study outcome is compared between the pre- and post-vaccination periods. In this design, it is important to consider the different control measures (or non-pharmaceutical interventions) imposed by governments to the population being analysed in these two periods. Some initiatives such as the Oxford Covid-19 Government Response Tracker – OxCGRT (43) systematically collects daily data on policy measures enforced by governments (e.g., school closures, travel restrictions, vaccination policy, lockdowns) to tackle COVID-19 since the beginning of the pandemic and across more than 180 countries, and define indicators which may help leverage the impact of a vaccination program taking into account the different stringency levels applied to the pre- and post-vaccination populations.
- *Cluster randomized vaccination trials.* This design is based on generating comparable social units called clusters by randomization. The outcome is compared between placebo and vaccine clusters. Cluster-randomized trials are usually conducted to quantify a treatment or intervention effect. In cluster-randomized trials, individuals are grouped based on specific characteristics (e.g., neighbourhood of residence), and the entire cluster is randomized to treatment or control. The process of randomization ensures that the treatment and control groups are exchangeable. This approach is useful when it is impractical or infeasible to randomize at the individual level. The randomized clusters can be compared to assess the overall impact of an intervention, which is particularly important in settings where intervention may have indirect effects (44).
- *Statistical modelling.* This design is normally associated with an outcome prediction (e.g., disease occurrence) without vaccination. It compares it to the population's occurrence with vaccination programs, henceforward named "epidemiological statistical modelling." This design can adjust for differences between populations, such as annual variations and secular disease trends or changes in health care use.

As shown in Table 1, none of the analysed articles followed the cluster-randomized vaccination trial study design. This is possibly due to the urge brought by the pandemic to vaccinate the worldwide population with vaccines which effectiveness has already been attested (45).

Nine out of the eighteen studies followed the epidemiological, statistical modelling study design, aiming to predict the impact on a community outcome by simulating scenarios with and without a vaccination roll-out. Meslé et al. (20), Galvani et al. (25), and Andrews et al. (26) estimated the number of either averted deaths or averted hospitalisations or both. To make these estimations possible, vaccine efficacy and effectiveness against deaths and hospitalisations studies were considered input variables of the impact study. In particular, Meslé et al. (20) proposed a standard approach to compare the estimated direct impact of the differential roll-out of COVID-19 vaccination programs across 33 countries in the WHO European Region, from December 2020 to November 2021. They calculated the weekly number of deaths averted per country taking the number of confirmed cases, vaccine coverage, and vaccine effectiveness in the given locality and time range into account, following Machado et al. (21). They also differed the vaccine coverage and effectiveness with at least one dose (which they called VU1 and VE1) from the vaccine coverage and effectiveness for those with complete vaccination schemas (VU2 and VE2), understanding that the number of vaccine doses influences the development of a full immune response individually, and consequently the protection from severe infection and death. Lower and upper bounds used for VE1 and VE2 were chosen based on observational studies for the vaccines most frequently used in the countries of that study. In their study, Meslé et al. (20) confirmed that both speed and extent of the vaccination in some eligible groups were determinants of vaccination impact with regard to averted deaths. Galvani et al. (25) also acknowledged the effectiveness of the different COVID-19 vaccine types administered in the US from October 2020 to June 2021 in preventing severe diseases, hospitalizations, and deaths due to COVID-19, which in turn contributed to increasing the impact of the vaccination program, potentially because of the vaccine's ability to reduce transmission of the virus.

The remaining studies followed the pre- and post-vaccination population comparison design. In McNamara et al. (22) and Rossman et al. (24), there were clear rules to define when a specific age group goes from pre-vaccination status to post-vaccination status. However, there is no such specification in Victora et al. (23), which analyses COVID-19 community outcomes over time, while vaccination coverage rises for the age groups studied.

Meslé et al. (20) applied the same formula to measure the averted deaths due to vaccination for all the populations from the 33 countries covered by their study. Even though all analyzed countries are from the WHO European region, they differ in many aspects, including geographical, sociodemographic, and vaccination programs, since each country applied vaccines from different manufacturers, which is even pointed out in the study. Moreover, the analysis described by Meslé et al. (20) also assumes that populations with and without vaccination programs (a.k.a. pre- and post-vaccination populations) have similar baseline transmission (hence the clustered populations are similar), which does not hold. Milman et al. (29) presented the relative change in the positive test fraction according to the change in the proportion of vaccinated individuals. Finally, data completeness is also essential for ecological studies. Complete and accurate data is fostered in the different health systems, but huge variation in quality and validity remains across organizations (46).

3.4. COVID-19 community outcomes analyzed

Although all articles address the impact of vaccination programs, the outcomes differ significantly. The results from the studies compare the dynamics of the pandemic based on different outcomes, with or without an ongoing vaccination program, and even simulating different vaccination scenarios. They calculate the variations in the disease outcome, which may refer to the reduction in cases, hospitalizations, deaths, or the number of deaths averted. It is important to note that a comparative analysis between studies is hampered by the different ways the impacts of vaccination processes are presented. Most studies estimated the impact in terms of averted COVID-19 deaths (20, 25, 26, 28, 32–34, 36, 37), and some of them also analyzed the averted hospitalizations (and severe hospitalizations) (24–26, 28, 34, 37). Other studies investigated the COVID-19

incidence seeking to estimate the reduction in the number of cases (19, 21, 28, 29, 34, 38).

There could be a more specific interest in low- and middle-income countries (LMICs) in view of the considerable obstacles in both receiving and distributing doses, especially at the beginning of the vaccination roll-out when vaccines were scarcer. In the three studies related to countries with disparities in access to healthcare and potential discrimination in vaccine distribution, the results are in line with those of developed countries in terms of impact. In both South American countries (Brazil and Colombia) and thirteen European countries (Albania, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Georgia, Republic of Moldova, Russian Federation, Serbia, North Macedonia, Turkey, and Ukraine), they successfully adopted strategies based on staggering vaccination in age groups, prioritizing older adults. All studies point to significant and relevant impacts of vaccine campaigns on the analyzed populations, whether due to the variation in the proportion of deaths in different age groups, the declines observed for the prioritized groups in the curves of cases and deaths, or the number of deaths avoided. The findings of each study are presented in Supplementary Table S1.

4. Discussion

The impact studies included in the present narrative review show significant differences in how they are developed and the main achieved outcomes. The analysis methods and tools are also quite different. We only selected articles based on actual vaccination data (even if combined with hypothetical vaccination scenarios) and those presented due to vaccine impact on the entire population. The selected studies covered European, Latin American, North American and Asian countries. The reviewed studies used data collected between December 2020 and June 2022.

Most COVID-19 vaccination campaigns worldwide have multiple vaccine platforms available to immunize a population. Therefore, the vaccination impact is not often associated with a vaccine from a single manufacturer. However, Israel exclusively used the Pfizer-BioNTech BNT162b2 mRNA COVID-19 vaccine. Thus, works from Rossman et al. (24), Milman et al. (29) and Haas et al. (28) could address the impact of a single platform vaccination campaign.

Cot et al. (19) established the relationship between the weekly percentage of the vaccinated population and the number of infections. The number of deaths averted by the vaccine is the main result of Meslé et al. (20), covering 33 European countries. Works from Rossman et al. (24), Andrews et al. (26), Galvani et al. (25), and Victora et al. (23) make use of temporal differences in the vaccination rate of different age groups to show a reduction in deaths, contamination, and/or hospitalizations for distinct age groups.

Regarding confounders, Rossman et al. (24), Andrews et al. (26), Galvani et al. (25), and Victora et al. (23) adjusted their results by age group. In Galvani et al. (25), the mobility rate was considered in the model. Notably, Victora et al. (23) and Galvani et al. (25) mentioned different variants of concern (VoCs) of the SARS-Cov virus; however, these VoCs should not be characterized as confounders of these studies since they were not explicitly taken into account in the models. Thus, pre- and post-vaccination populations were assumed to have similar baseline transmission. These studies only mentioned the VoCs that were dominant in the studied populations: Victora et al. (23) study was conducted when gamma was the dominant VoC, while Galvani et al. (25) was conducted during the dominance of the Alpha, Gamma, Delta, and the original Wuhan-1 variants. Likewise, vaccine manufacturers were not explicitly addressed in the models to calculate the impact and should not be considered confounder variables.

The studies also differ in outcomes, involving deaths, hospital admissions, incidences, non-ICU hospitalizations, ICU hospitalizations, and symptomatic cases. Machado et al. (27), Cot et al. (19), and Rossman et al. (24) addressed the impact of other interventions or occurrences used as parameters; the following interventions or occurrences were mentioned: pre-existing immunity, self-isolation of infected individuals, state stay-home order, state facemask police or proportions of members of public who wear masks in public and, finally, lockdowns. Rossman et al. (24) and Machado et al. (27) used the impact of lockdown as a model parameter.

The way of presenting the results is also quite different. Most articles present the number of cases, hospitalizations or deaths averted. In some studies, the asymmetry in the temporality of the vaccination process between different age groups is used to point out how it affects the relative participation of these groups in the total number of cases, deaths or hospitalizations. There are approaches that establish comparisons between countries and territories. In these cases, the different vaccination rates observed are related to different declines in the numbers of cases, hospitalizations or deaths. There is a specific article that studies the differentiated impact determined by the different intervals between doses. Finally, there is a study in which the authors identify the minimum weekly vaccination rate to guarantee a specific value for the number of deaths.

Several of the reviewed articles made use of epidemiological models such as SEIR, which can be seen as simplistic, that is, using a few compartments instead of thousands of compartments to represent the real complexity of the system. In an epidemic, several phenomena are difficult to understand, caused by the interaction of a huge number of agents which, even when acting locally, are capable of influencing results elsewhere (47). However, SEIR models proved robust enough to be applied in different geographic locations and in populations of different ethnic origins, enabling and recommending their use in territories with a lack of viral testing. Barbosa et al. (48) applied the SEIR model using epidemiological data from Marabá, a poor municipality in the state of Pará in Brazil, with estimated values of latency time and infectious time obtained in Chinese populations and these proved to be useful for predicting the evolution of COVID-19 cases, a more complex process than the estimation of the vaccine impact. One last important consideration concerning the presented models is that they do not recognize the prolonged duration of the pandemic and the representative rate of deaths during a given period, since they do not use the number of deaths and births in the most active period of the pandemic as parameters, thus, disregarding its vital dynamics, foreseen in the complex systems applied to epidemiological models presented by Lima (47).

Figure 2 synthesizes the main characteristics of the vaccination impact studies analyzed in this narrative review. Although the lack of systematicity in the review process does not allow the complete range of designs, methods, variables, and results of vaccine impact studies to be presented, it is understood that the outline of its characteristics is broad enough to give those interested in the subject the breadth of possibilities available.

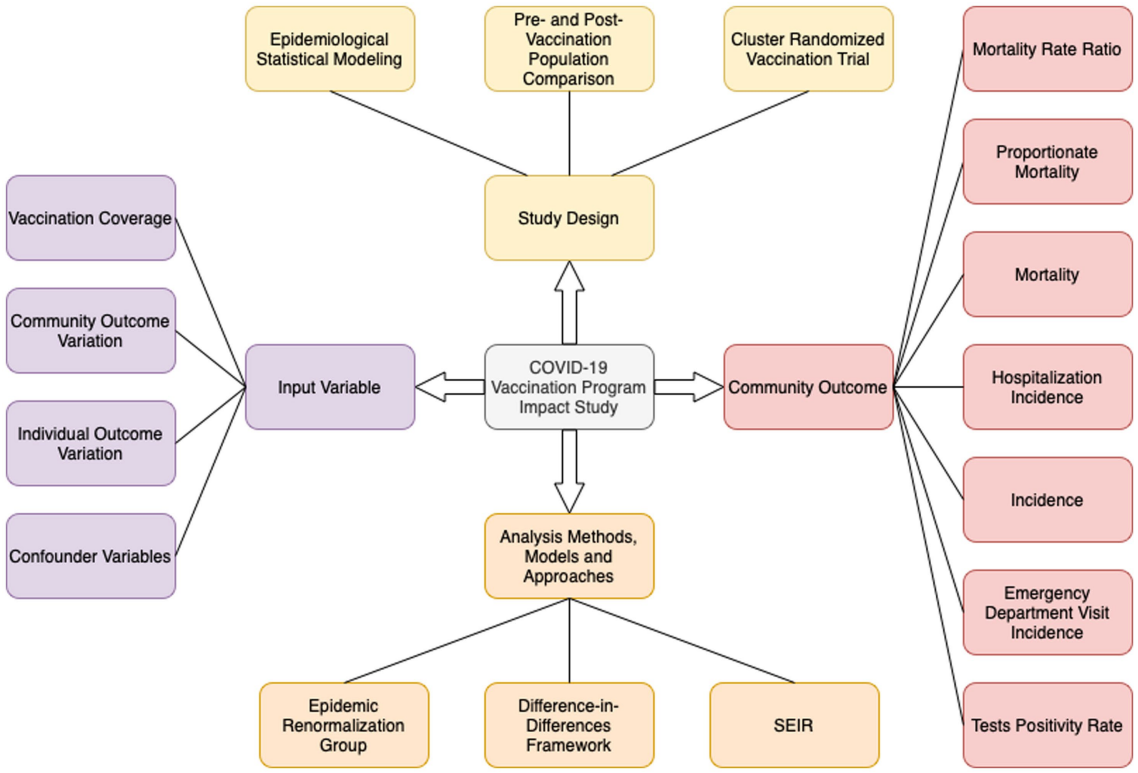


FIGURE 2
The main characteristics of the vaccination impact studies analysed in this review. As visually represented, a vaccination impact study is distinguished by four perspectives: Input variables, Community outcome, Study design, and Analysis methods, models, and approaches.

5. Conclusion

The articles analyzed in this Narrative Review, regardless of the methods applied and country(ies) covered, share in their results the significant population impact brought by the vaccination process. Although the pandemic is cooling down at the moment, its permanence has required new booster doses to be administered. In addition, the possibilities for the emergence of new variants of concern can alter vaccine efficacy, establishing new levels of vaccine impact. Studying the impact of COVID-19 vaccines will remain the slogan of the day for some time. The present work contributed to the research on this theme, offering a broad and structured view of the methodological possibilities, models, approaches and designs. Furthermore, it aims to contribute to a broader view of the possible studies, as it also brings together the different possibilities of input variables adopted and different outcome variables that may represent the vaccination impact.

There is, however, an approach to vaccine impact that remains underexplored. In addition to disparities in the application of COVID-19 tests and in the supply of protective equipment, LMICs suffered from problems related to the availability of vaccines (49). In none of the analyzed studies, the discussion on the superiority of the strategy, adopted by the richest countries, in terms of vaccine impact, of protecting them-selves other than globally controlling the COVID-19 pandemic, was privileged. In one of the studies, Louden (50) says that careful consideration of vaccine production, pricing, allocation, and distribution must be taken into account to ensure equitable access to COVID-19 vaccines scaling up the global COVID-19 vaccination program but in this study vaccine impact was not the approach. Ali et al. (51) discussed the problem of vaccine equity in LMICs. They found that inequalities in wealth, education, and geographic access can affect vaccine impact and vaccination dropout which demands more attention in countries where the level of inequality is considerably higher. The analysis of global vaccination rollouts comparing LMIC to rich countries should include each country demography (and the age groups approved to be vaccinated). LMICs with low proportion of population older than 60 years cannot be direct compared to some European countries with high proportion of population older than 60 years. Furthermore, the propensity of young adults to get vaccinated in a country with a young population is different than in one with an old population. The assessment of the benefits of potentially protecting older adults to the risks of the vaccine is dependent on demography.

Another study alternative, which was not observed, would be to evaluate the vaccine impact in terms of the relative dynamics of cases and deaths after the consolidation of the vaccination process, comparing the results between countries before and after vaccination, as a function of the percentage of vaccinated and the number of doses administered. Brazil, for example, a LMIC, due to the greater adherence of its population to vaccination, after being the eighth country in deaths by COVID-19 in the world in the period of 2020–2021, already occupies the 17th position in October of 2022, considering the entire pandemic period, registering fewer deaths *per capita*, bearing in mind only the year 2022, than the United States and many European developed countries. Finally, the impact of vaccination campaigns could be analyzed in terms of compliance with COVID-19 regulations, mobility and contact behavior in communities, the likelihood of transmission at a potentially infectious contact, human behavior (and mental health), and health care costs.

Author contributions

HF, LM, FBa, and IP: conceptualization, formal analysis, investigation, and visualization. HF, LM, FBa, and IP: methodology and writing—original draft preparation. FBa, IP, OR, SH, PM, and FBo: validation. HF, LM, and IP: data curation. IP, OR, SH, PM, and FBa: writing—review and editing. OR, SH, and FBo: supervision. All authors contributed to the article and approved the submitted version.

Funding

This work was performed as part of the Grand Challenges ICODA pilot initiative (INV-017293), funded by the Minderoo Foundation and the Bill & Melinda Gates Foundation. The research was also supported by the National Council for Scientific and Technological Development (CNPq) [grant numbers 310940/2019-2 and 403863/2016-3 to SH; 422810/2021-5 and 312059/2022-1 to FBa; and 422470/2021-0 and 311519/2022-9 to PM], Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro (FAPERJ) [grant numbers 211.308/2019 to FBa; and 211.086/2019, 211.645/2021, and 201.348/2022 to PM], and the Pontifical Catholic University of Rio de Janeiro (PUC-Rio). OTR is funded by Sara Borrell from Instituto de Salud Carlos III (CD19/00110).

Acknowledgments

This research was conducted in the Center for Healthcare Operations and Intelligence - NOIS (www.nois.ind.puc-rio.br) in PUC-Rio. The authors thank all NOIS participants for their collaborative production of scientific analyses of the COVID-19 pandemic in Brazil. OTR acknowledges the support from the Generalitat de Catalunya through the CERCA Program and from the Spanish Ministry of Science and Innovation and State Research Agency through the "Centro de Excelencia Severo Ochoa 2019-2023" Program (CEX2018- 000806-S).

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1126461/full#supplementary-material

References

1. Ortiz JR, Neuzil KM. The value of vaccine programme impact monitoring during the COVID-19 pandemic. *Lancet*. (2022) 399:119–21. doi: 10.1016/S0140-6736(21)02322-9

2. Bernal JL, Andrews N, Gower C, Gallagher E, Simmons R, Thelwall S, et al. Effectiveness of Covid-19 vaccines against the B.1.617.2 (Delta) variant. *N Engl J Med*. (2021) 385:585–94. doi: 10.1056/NEJMoa2108891

3. Jara A, Undurraga EA, González C, Paredes F, Fontecilla T, Jara G, et al. Effectiveness of an inactivated SARS-CoV-2 vaccine in Chile. *N Engl J Med*. (2021) 385:875–84. doi: 10.1056/NEJMoa2107715

4. Hunter PR, Brainard J (2021). Estimating the effectiveness of the Pfizer COVID-19 BNT162b2 vaccine after a single dose. A reanalysis of a study of 'real-world' vaccination outcomes from Israel. medRxiv [Preprint]. doi: 10.1101/2021.02.01.21250957

5. World health Organization (2021). Evaluation of COVID-19 vaccine effectiveness: interim guidance. Available at: https://www.who.int/publications-detail-redirect/WHO-2019-nCoV-vaccine_effectiveness-measurement-2021.1 (Accessed April 11, 2022).

6. Pritchard E, Matthews PC, Stoesser N, Eyre DW, Gethings O, Vihta K-D, et al. Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom. *Nat Med*. (2021) 27:1370–8. doi: 10.1038/s41591-021-01410-w

7. Tande AJ, Pollock BD, Shah ND, Farrugia G, Virk A, Swift M, et al. Impact of the coronavirus disease 2019 (COVID-19) vaccine on asymptomatic infection among patients undergoing Preprocedural COVID-19 molecular screening. *Clin Infect Dis*. (2022) 74:59–65. doi: 10.1093/cid/ciab229

8. Moghadas SM, Vilches TN, Zhang K, Wells CR, Shoukat A, Singer BH, et al. The impact of vaccination on coronavirus disease 2019 (COVID-19) outbreaks in the United States. *Clin Infect Dis*. (2021) 73:2257–64. doi: 10.1093/cid/ciab079

9. Pasquale S, Gregorio GL, Caterina A, Francesco C, Beatrice PM, Vincenzo P, et al. COVID-19 in low- and middle-income countries (LMICs): A narrative review from prevention to vaccination strategy. *Vaccine*. (2021) 9:1477. doi: 10.3390/vaccines9121477

10. Eroglu B, Nuwarda RF, Ramzan I, Kayser V. A narrative review of COVID-19 vaccines. *Vaccines*. (2022) 10:62. doi: 10.3390/vaccines10010062

11. Hafiz I, Illian DN, Meila O, Utomo ARH, Susilowati A, Susetya IE, et al. Effectiveness and efficacy of vaccine on mutated SARS-CoV-2 virus and post vaccination surveillance: a narrative review. *Vaccine*. (2022) 10:82. doi: 10.3390/vaccines10010082

12. Steffens MS, Bullivant B, Bolsewicz K, King C, Beard F. Factors influencing COVID-19 vaccine acceptance in high income countries prior to vaccine approval and rollout: a narrative review. *Int J Public Health*. (2022) 67:1604221. doi: 10.3389/ijph.2022.1604221

13. Saadi N, Chi Y-L, Ghosh S, Eggo RM, McCarthy CV, Quaife M, et al. Models of COVID-19 vaccine prioritisation: a systematic literature search and narrative review. *BMC Med*. (2021) 19:318. doi: 10.1186/s12916-021-02190-3

14. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. (2009) 339:b2700. doi: 10.1136/bmj.b2700

15. Moher D, Liberati A, Tetzlaff J, Altman DGGroup TP. Preferred reporting items for systematic reviews and Meta-analyses: the PRISMA statement. *PLoS Med*. (2009) 6:e1000097. doi: 10.1371/journal.pmed.1000097

16. Thomé AMT, Scavarda LF, Scavarda AJ. Conducting systematic literature review in operations management. *Prod Plan Control*. (2016) 27:408–20. doi: 10.1080/09537287.2015.1129464

17. Peres IT, Hamacher S, Oliveira FLC, Thomé AMT, Bozza FA. What factors predict length of stay in the intensive care unit? Systematic review and meta-analysis. *J Crit Care*. (2020) 60:183–94. doi: 10.1016/j.jcrc.2020.08.003

18. Wohlin C. (2014). "Guidelines for snowballing in systematic literature studies and a replication in software engineering." in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*. London England United Kingdom: ACM. pp. 1–10.

19. Cot C, Cacciapaglia G, Islind AS, Óskarsdóttir M, Sannino F. Impact of US vaccination strategy on COVID-19 wave dynamics. *Sci Rep*. (2021) 11:10960. doi: 10.1038/s41598-021-90539-2

20. Meslé MM, Brown J, Mook P, Hagan J, Pastore R, Bundle N, et al. Estimated number of deaths directly averted in people 60 years and older as a result of COVID-19

vaccination in the WHO European region, December 2020 to November 2021. *Eur Secur*. (2021) 26:2101021. doi: 10.2807/1560-7917.ES.2021.26.47.2101021

21. Machado A, Mazagatos C, Dijkstra F, Kislaya I, Gherasim A, McDonald SA, et al. Impact of influenza vaccination programmes among the elderly population on primary care, Portugal, Spain and the Netherlands: 2015/16 to 2017/18 influenza seasons. *Eur Secur*. (2019) 24:1900268. doi: 10.2807/1560-7917.ES.2019.24.45.1900268

22. McNamara LA, Wiegand RE, Burke RM, Sharma AJ, Sheppard M, Adjemian J, et al. Estimating the early impact of the US COVID-19 vaccination programme on COVID-19 cases, emergency department visits, hospital admissions, and deaths among adults aged 65 years and older: an ecological analysis of national surveillance data. *Lancet*. (2022) 399:152–60. doi: 10.1016/S0140-6736(21)02226-1

23. Victora CG, Castro MC, Gurzenda S, Medeiros AC, França GVA, Barros AJD. Estimating the early impact of vaccination against COVID-19 on deaths among elderly people in Brazil: analyses of routinely-collected data on vaccine coverage and mortality. *eClinicalMedicine*. (2021) 38:101036. doi: 10.1016/j.eclinm.2021.101036

24. Rossman H, Shilo S, Meir T, Gorfine M, Shalit U, Segal E. COVID-19 dynamics after a national immunization program in Israel. *Nat Med*. (2021) 27:1055–61. doi: 10.1038/s41591-021-01337-2

25. Galvani A, Moghadas SM, Schneider EC (2021). Deaths and hospitalizations averted by rapid U.S. vaccination rollout. The Commonwealth Fund. Issue Briefs. doi: 10.26099/wm2j-mz32

26. Andrews N, Stowe J, Ismael S. Impact of COVID-19 vaccines on mortality in England. *Public Health Engl*. (2021). Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/977249/PHE_COVID-19_vaccine_impact_on_mortality_March.pdf

27. Machado B, Antunes L, Caetano C, Pereira JF, Nunes B, Patrício P, et al. The impact of vaccination on the evolution of COVID-19 in Portugal. *Math Biosci Eng MBE*. (2022) 19:936–52. doi: 10.3934/mbe.2022043

28. Haas EJ, McLaughlin JM, Khan F, Angulo FJ, Anis E, Lipsitch M, et al. Infections, hospitalisations, and deaths averted via a nationwide vaccination campaign using the Pfizer–BioNTech BNT162b2 mRNA COVID-19 vaccine in Israel: a retrospective surveillance study. *Lancet Infect Dis*. (2022) 22:357–66. doi: 10.1016/S1473-3099(21)00566-1

29. Milman O, Yelin I, Aharony N, Katz R, Herzel E, Ben-Tov A, et al. Community-level evidence for SARS-CoV-2 vaccine protection of unvaccinated individuals. *Nat Med*. (2021) 27:1367–9. doi: 10.1038/s41591-021-01407-5

30. Miłobędzki P. Are vaccinations alone enough to curb the dynamics of the COVID-19 pandemic in the European Union? *Econometrics*. (2022) 10:25. doi: 10.3390/econometrics10020025

31. Liu Y, Pearson CAB, Sandmann FG, Barnard RC, Kim J-H, Flasche S, et al. Dosing interval strategies for two-dose COVID-19 vaccination in 13 middle-income countries of Europe: health impact modelling and benefit-risk analysis. *Lancet Reg Health – Eur*. (2022) 17:100381. doi: 10.1016/j.lanepe.2022.100381

32. Caetano C, Morgado ML, Patrício P, Leite A, Machado A, Torres A, et al. Measuring the impact of COVID-19 vaccination and immunity waning: a modelling study for Portugal. *Vaccine*. (2022) 40:7115–21. doi: 10.1016/j.vaccine.2022.10.007

33. Rojas-Botero ML, Fernández-Niño JA, Arregocés-Castillo L, Ruiz-Gómez F. Estimated number of deaths directly avoided because of COVID-19 vaccination among older adults in Colombia in 2021: an ecological, longitudinal observational study. *F1000Research*. (2022) 11:198. doi: 10.12688/f1000research.109331.3

34. Sacco C, Mateo-Urdiales A, Petrone D, Spuri M, Fabiani M, Vescio MF, et al. Estimating averted COVID-19 cases, hospitalisations, intensive care unit admissions and deaths by COVID-19 vaccination, Italy, January–September 2021. *Eur Secur*. (2021) 26:2101001. doi: 10.2807/1560-7917.ES.2021.26.47.2101001

35. Bonmarin I, Belchior E, Lévy-Bruhl D. Impact of influenza vaccination on mortality in the French elderly population during the 2000–2009 period. *Vaccine*. (2015) 33:1099–101. doi: 10.1016/j.vaccine.2015.01.023

36. Mattiuzzi C. (2021). COVID-19 vaccination uptake strongly predicts averted deaths of older people across Europe (preprint).

37. Shoukat A, Vilches TN, Moghadas SM, Sah P, Schneider EC, Shaff J, et al. Lives saved and hospitalizations averted by COVID-19 vaccination in new York City: a modeling study. *Lancet Reg Health Am*. (2022) 5:100085. doi: 10.1016/j.lana.2021.100085

38. Suthar AB, Wang J, Seffren V, Wiegand RE, Griffing S, Zell E. Public health impact of covid-19 vaccines in the US: observational study. *BMJ*. (2022) 377:e069317. doi: 10.1136/bmj-2021-069317

39. Iboi EA, Ngonghala CN, Gumel AB. Will an imperfect vaccine curtail the COVID-19 pandemic in the US? *Infect Dis Model*. (2020) 5:510–24. doi: 10.1016/j.idm.2020.07.006

40. Fang F, Clemens JD, Zhang Z-F, Brewer TF. (2021). Impact of SARS-CoV-2 Vaccines on Covid-19 incidence and mortality in the United States. medRxiv [Preprint]. 2021.11.16.21266360. doi: 10.1101/2021.11.16.21266360

41. Levin KA. Study design VI - ecological studies. *Evid Based Dent*. (2006) 7:108–8. doi: 10.1038/sj.ebd.6400454

42. Hanquet G, Valenciano M, Simondon F, Moren A. Vaccine effects and impact of vaccination programmes in post-licensure studies. *Vaccine*. (2013) 31:5634–42. doi: 10.1016/j.vaccine.2013.07.006

43. Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, Phillips T, et al. A global panel database of pandemic policies (Oxford COVID-19 government response tracker). *Nat Hum Behav*. (2021) 5:529–38. doi: 10.1038/s41562-021-01079-8

44. Kilpatrick KW, Hudgens MG, Halloran ME. Estimands and inference in cluster-randomized vaccine trials. *Pharm Stat*. (2020) 19:710–9. doi: 10.1002/pst.2026

45. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*. (2015) 350:h391. doi: 10.1136/bmj.h391

46. Saunders C, Abel G. Ecological studies: use with caution. *Br J Gen Pract*. (2014) 64:65–6. doi: 10.3399/bjgp14X676979

47. de Oliveira Lima JP. Sistemas Complexos aplicado a modelos epidemiológicos. *Phys Organum - Rev Estud Física UnB*. (2021) 7:59–71.

48. Barbosa WF, da Costa Moreira EB, de Araújo JM, Pazin-Filho A, CDF B. Modelo SEIR para avaliação do comportamento da pandemia de Covid-19 em Marabá-PA. *Rev Med*. (2021) 100:322–8. doi: 10.11606/issn.1679-9836.v100i4p322-328

49. Park M-B, Ranabhat CL. COVID-19 trends, public restrictions policies and vaccination status by economic ranking of countries: a longitudinal study from 110 countries. *Arch Public Health*. (2022) 80:197. doi: 10.1186/s13690-022-00936-w

50. Louden EM. Scaling up the global COVID-19 vaccination program: production, allocation, and distribution with an emphasis on equity. *Yale J Biol Med*. (2022) 95:379–87.

51. Ali HA, Hartner A-M, Echeverria-Londono S, Roth J, Li X, Abbas K, et al. Vaccine equity in low and middle income countries: a systematic review and meta-analysis. *Int J Equity Health*. (2022) 21:82. doi: 10.1186/s12939-022-01678-5

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 02 June 2023
DOI 10.3389/fpubh.2023.1150810

OPEN ACCESS

EDITED BY
Pierpaolo Ferrante,
National Institute for Insurance against Accidents at Work (INAIL), Italy

REVIEWED BY
Thomas Finnie,
UK Health Security Agency (UKHSA), United Kingdom
Bernard Joseph Hudson,
New South Wales Health Pathology, Australia

*CORRESPONDENCE
Nick Scott
✉ Nick.Scott@burnet.edu.au

RECEIVED 25 January 2023
ACCEPTED 05 May 2023
PUBLISHED 02 June 2023

CITATION
Abeysuriya RG, Sacks-Davis R, Heath K, Delport D, Russell FM, Danchin M, Hellard M, McVernon J and Scott N (2023) Keeping kids in school: modelling school-based testing and quarantine strategies during the COVID-19 pandemic in Australia.
Front. Public Health 11:1150810.
doi: 10.3389/fpubh.2023.1150810

Keeping kids in school: modelling school-based testing and quarantine strategies during the COVID-19 pandemic in Australia

Romesh G. Abeysuriya[1,2], Rachel Sacks-Davis[1,2], Katherine Heath[1], Dominic Delport[1], Fiona M. Russell[3,4], Margie Danchin[3,4,5], Margaret Hellard[1,2,6,7,8], Jodie McVernon[8,9] and Nick Scott[1,2]*

[1]Disease Elimination Program, Burnet Institute, Melbourne, VIC, Australia, [2]Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, VIC, Australia, [3]Murdoch Children's Research Institute, Parkville, VIC, Australia, [4]Department of Paediatrics, The University of Melbourne, Parkville, VIC, Australia, [5]The Royal Children's Hospital, Melbourne, VIC, Australia, [6]School of Population and Global Health, The University of Melbourne, Parkville, VIC, Australia, [7]Department of Infectious Diseases, The Alfred and Monash University, Melbourne, VIC, Australia, [8]Department of Infectious Diseases, The University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia, [9]Victorian Infectious Diseases Epidemiology Unit, The Royal Melbourne Hospital at the Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia

Background: In 2021, the Australian Government Department of Health commissioned a consortium of modelling groups to generate evidence assisting the transition from a goal of no community COVID-19 transmission to 'living with COVID-19', with adverse health and social consequences limited by vaccination and other measures. Due to the extended school closures over 2020–21, maximizing face-to-face teaching was a major objective during this transition. The consortium was tasked with informing school surveillance and contact management strategies to minimize infections and support this goal.

Methods: Outcomes considered were infections and days of face-to-face teaching lost in the 45 days following an outbreak within an otherwise COVID-naïve school setting. A stochastic agent-based model of COVID-19 transmission was used to evaluate a 'test-to-stay' strategy using daily rapid antigen tests (RATs) for close contacts of a case for 7 days compared with home quarantine; and an asymptomatic surveillance strategy involving twice-weekly screening of all students and/or teachers using RATs.

Findings: Test-to-stay had similar effectiveness for reducing school infections as extended home quarantine, without the associated days of face-to-face teaching lost. Asymptomatic screening was beneficial in reducing both infections and days of face-to-face teaching lost and was most beneficial when community prevalence was high.

Interpretation: Use of RATs in school settings for surveillance and contact management can help to maximize face-to-face teaching and minimize outbreaks. This evidence supported the implementation of surveillance testing in schools in several Australian jurisdictions from January 2022.

KEYWORDS

COVID-19, outbreak, rapid antigen test, agent-based model, school, surveillance, quarantine

Introduction

Until mid-2021, Australia endeavored to strongly suppress community SARS-CoV-2 transmission by limiting incursions with tight border controls and containing outbreaks with contact tracing and strict community restrictions, including extended lockdowns in some jurisdictions. However, the rollout of COVID-19 vaccines meant that Australia could consider alternate approaches to managing COVID-19 without either the social and economic impact associated with containing outbreaks, or the dire health outcomes associated with community transmission prior to vaccines being available. In late 2021, the Commonwealth Government commissioned the Doherty Institute to lead a consortium of modelling groups to support development of the National Plan to transition Australia's COVID-19 response (1).

While Australia pursued a goal of zero community SARS-CoV-2 transmission (2020–2021), schools were often closed as part of broader lockdown measures. When schools were open, outbreaks in most jurisdictions were managed by reactive closures and targeted periods of quarantine. This typically involved a school closing for cleaning following a positive case (often two to 3 days), followed by a 14-day quarantine for all close contacts and their households. Specific school outbreak strategies differed between states and territories but were generally commensurate with other community restrictions, and were successful in reducing the size of school outbreaks (2). However, due to concern about the potential health and psychosocial impacts of these public health responses on children (3), maximizing in-person learning was seen as a national priority.

With Australia's transition to "living with COVID-19" and increased community transmission, the rates of incursions into schools were anticipated to be higher than previously experienced (4). In this context, a quarantine-based approach in schools was recognized to be unsustainable and inconsistent with the national priority of maximizing face-to-face teaching. Hence there was a need for different approaches to managing cases and contacts in schools and keeping schools open safely.

This commissioned work was developed through a participatory process with the Commonwealth Departments of Health and Education. Its agreed focus was to support face-to-face teaching through identification of strategies that would minimize importation and transmission of infections in the school environment. Given this objective, infections and days of face-to-face teaching lost following an imported case were defined as the key outcomes of interest. The specific aims of this study were to evaluate the potential impacts of two strategies identified in consultation: the use of rapid antigen tests (RATs) for asymptomatic screening, or as an alternative risk mitigation to home quarantine of close contacts, based on their effective implementation in other country settings (5–7). The scope of enquiry was restricted to transmission risks anticipated in primary and secondary day schools, and analyses were conducted prior to the emergence of the Omicron variant.

Methods

Model overview

An established agent-based microsimulation model, *Covasim* (8), was used to simulate outbreaks in school settings. The model is open-source and available online (9) and has previously been used to model epidemic waves and response strategies in Australia (10–12). Additional model details are provided in the Supplementary material. The code for the simulations and analysis presented in this study is available at https://github.com/optimamodel/covasim_aus_schools.

To address the specific questions of this study, we implemented a more detailed model of school contact networks than those provided by *Covasim*, based on the structure of the Australian school system. Aside from differences in age and vaccine eligibility, there are important differences in social and mixing structure between primary and secondary schools; hence these two settings were modelled separately. Schools were characterized by three types of interaction – transmission within classrooms involving students and teachers, transmission outside of classrooms between students (e.g., during breaks), and transmission outside of classrooms between teachers (e.g., in staff common rooms).

Primary schools

In Australia, primary school students are typically assigned to a single class for all lessons. Primary schools were therefore modelled as a collection of classrooms, aggregated into schools. To construct the primary school contact network, children from ages 5–11 were assigned to schools. Within each school, students were assigned to classrooms with others of the same age, and each classroom had an assigned teacher (randomly selected from the working-age population, age 18–65). Each classroom was a fully connected cluster, with transmission possible between any pair of individuals within the same classroom. The classroom contact network was therefore highly clustered (Figure 1A). Non-classroom mixing was incorporated by assigning each student additional contacts randomly selected from the entire school.

Secondary schools

Secondary school students in Australia are typically grouped into separate classes for each subject and have around 4–6 classes per day. To simplify the implementation and parameterization of the model, instead of modelling each classroom explicitly, students were assigned a number of random contacts amongst other students of the same age, with the number of contacts reflecting the average number of different students they would encounter each day. Each student was then randomly assigned a number of contacts with staff members reflecting the typical number of classes per day, and a number of contacts with other students randomly selected from the entire school to account for non-classroom mixing. Compared to primary schools, there is considerably more mixing between students and teachers within classrooms, and classroom contacts are much less clustered (Figure 1B).

Disease transmission

Transmission in the model has a probability of occurring each time a susceptible individual is in contact with an infectious individual through one of their contact networks. The overall transmission

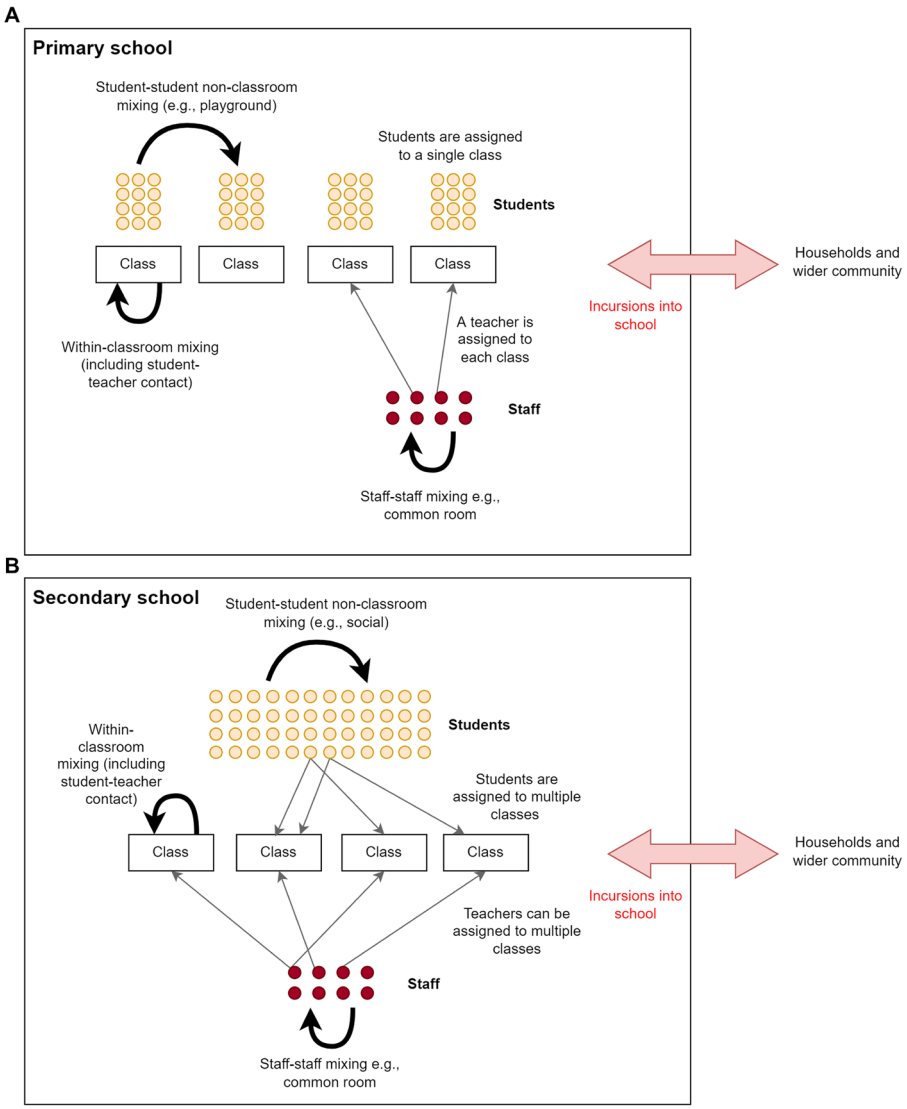


FIGURE 1
Contact networks within schools in the model for a) primary schools, and b) secondary schools. Schools included student–student classroom contacts, student–student non-classroom contacts, teacher-teacher contacts and teacher-student contacts. Primary schools were modelled as a collection of classrooms, where students of the same age are assigned a teacher. Secondary school students and teachers have more contacts than primary school students and teachers because they attend multiple classes.

probability per contact per day was calibrated based on the Delta variant epidemic wave in Melbourne over the July–September 2021 period (12). For individual contacts, this transmission risk was further weighted according to the setting of the contact (e.g., classroom, home), the time-varying viral load of the person infected, whether or not they have symptoms (based on an age-specific probability of being symptomatic), and an age-specific disease susceptibility (Table 1).

Symptomatic testing probability (COVID-19 cases)

All people with severe disease were assumed to be tested. For people with mild symptoms, the model included a per-day probability of seeking a test, which was necessary for the first case to be diagnosed when surveillance was not in place (noting that the first case to be detected may be a household member of a student at the school, which would trigger contact tracing for the student). The symptomatic testing model calibration process estimated that people with mild symptoms who were not identified through contact tracing would seek testing during their symptomatic period with a per-day testing probability of 11% (varied in a sensitivity analysis).

Vaccination

All baseline scenarios were run with 80% COVID-19 vaccination coverage in ages 12+ and 100% vaccination coverage in teachers, reflecting likely uptake of the vaccine at the time of analysis, the likelihood of vaccines becoming mandatory for teachers, and the fact that vaccines for ages 5–11 were yet to be approved for use in Australia at the time of reporting. However, sensitivity analyses were

TABLE 1 Model parameters related to schools.

Parameter area	Estimate	Source
Primary school		
Average school size	298	Number of primary students (2,268,891 full time + part time in 2020; ABS (13) Table 42b) divided by number of Primary + Primary/secondary schools (6249 + 1363 in 2020; ABS (13) Table 35b).
Average class size	22	Average class size of primary schools. Victorian government (14), with class sizes sampled from their distribution in analyses.
Average number of student–student non-classroom contacts per day, per student	2	Assumption; tested in sensitivity analysis. This impacts the efficacy of test-to-stay of class contacts verses close contacts or entire school.
Average number of teacher-teacher contacts per day, per teacher	5	Assumption.
Vaccination coverage	0%	Vaccines for under-12 children were not authorized at the time of analysis (a sensitivity analysis including vaccination in primary school children is provided in the **Supplementary material**).
Secondary school		
Average students per school	622	Number of secondary students (1,738,083 full time + part time in 2020; ABS (13) Table 42b) divided by number of Secondary + Primary/secondary schools (1433 + 1363 in 2020; ABS (13) Table 35b)
Average teacher/student ratio	12	ABS data (13). Suggesting secondary schools have on average 12.1 students to one teacher.
Average number of student–student classroom contacts per day	44	Average class size in secondary school of 22 ((15); page 354), assuming two unique classrooms of contacts per student per day.
Average number of student–student non-classroom contacts per day	5	Assumption; tested in sensitivity analysis. This impacts the efficacy of test-to-stay of class contacts verses close contacts or entire school.
Average number of teacher-teacher contacts per day	5	Assumption.
Average number of teacher-student contacts per day, per student	6	Assumes students have six classes per day
Vaccine coverage	80%	Assumed peak coverage level, based on expected vaccine uptake at the time of analysis.
Probability of transmission per contact per day (without vaccines or NPIs)		
Student–student (primary classroom)	0.25	Delphi process; Scott et al. (10) Measured as relative to household transmission per contact - e.g. a typical day's worth of contact in school is 75% less likely to result in transmission than a typical day's worth of contact at home. Non-classroom primary school contacts equivalent to outdoor contacts; secondary school classroom contacts halved to account for shorter interactions. Note: these are not attack rates, and all transmission probabilities are scaled by an overall calibration parameter, as well as age-specific susceptibility, vaccine status, and NPIs in place. Attack rates also depend on frequency and number of contacts. All transmission probabilities were varied in sensitivity analyses when NPI efficacy is tested.
Student–student (primary non-classroom)	0.03	
Student–student (secondary class contact)	0.12	
Student–student (secondary close/social contact)	0.12	
Teacher-teacher	0.25	Assumption that transmission risks in schools are equivalent for all types of contacts. Note that the model has independent parameters to account for differences in susceptibility by age
Teacher-student (primary)	0.25	
Teacher-student (secondary)	0.12	
Age-susceptibility (relative to 20–49year old)		
Age 0–4	0.349	Derived from Davies et al. (16)
Age 5–9	0.423	
Age 10–14	0.495	
Age 15–19	0.599	
Age 20–24	0.846	
Age 24–29	1	
Probability of being symptomatic		
Age 0–9	0.28	Davies et al. (16)
Age 10–19	0.20	
Age 20–29	0.26	
Rapid antigen testing (RAT)		
Sensitivity	0.773	Muhi et al. (17) Lower bound selected to account for inconsistent self-use. Note that PCR is modelled as having 87% sensitivity in real world use (systematic review Arevalo-Rodriguez et al. (18))

run to investigate the extent to which secondary schools in the baseline scenarios already benefited from vaccination and to assess the benefits of vaccines for primary school students if they became available (provided in the Supplementary material). Vaccine parameters were based on efficacy estimates against the Delta variant available at the time of analysis (19, 20) (see Supplementary material).

School testing and tracing strategies

Two overarching strategies for incorporating RATs into schools were considered. The first was asymptomatic surveillance, where students were required to take RATs regularly regardless of any symptoms. The specific implementations considered were: no surveillance testing; twice weekly teacher testing with RATs; and twice weekly student testing with RATs. These scenarios were considered with and without contact tracing in place.

The second strategy considered was a "test-to-stay" scheme in which contacts of diagnosed cases performed daily RATs instead of being quarantined. The specific implementations of the test-to-stay scheme were: no contact tracing (neither testing nor isolation for contacts); 7 days quarantine of classroom contacts with/without daily RAT; daily RAT of classroom contacts who are permitted to remain at school so long as they test negative ("test-to-stay"); and entire school test-to-stay with daily RAT after initial case detection.

The scenario with both seven-day quarantine and daily testing for 7 days enables assessment of the benefit of quarantine incremental to test-to-stay, controlling for differences in case ascertainment. Contact tracing scenarios were based around classroom contacts rather than all contacts, as classroom contacts were deemed more practical to identify in response actions. Scenarios were also examined with both asymptomatic surveillance and test-to-stay to assess the incremental impact of surveillance strategies when combined with contact tracing.

In all scenarios, students or teachers diagnosed with COVID-19 were assumed to be removed from the school and required to isolate until no longer infectious.

Model simulations and outcomes

The model was initialized with a single infection allocated randomly within a school. The model was then run for 45 days to allow sufficient time for the outbreak to grow within the school, while limiting the impact of broader community transmission on within-school dynamics. The number of cumulative infections in students or teachers attending the school were recorded. Infections were used as the primary outcome measure as opposed to diagnosed cases to avoid bias when comparing strategies with different testing rates. Importantly, although we report outcomes after 45 days, in cases where there is substantial transmission the outbreak is likely to be ongoing, and further transmission would take place after the simulation timeframe. In these cases, the cumulative number of infections after 45 days serves as a proxy measure for the growth rate of the outbreak. We elected not report the basic reproduction number R_0 because the short time window, small size, and wide range of stochastic outcomes makes this metric difficult to interpret for the outbreaks modelled in this study.

For each scenario, the simulation was repeated 1000 times and reported outcomes are based on the distributions of (1) secondary infections occurring in the same school; and (2) days of face-to-face teaching lost. Days of face-to-face teaching lost are calculated for the school as the total student-days spent in isolation or quarantine as a result of a school's testing and quarantine policy over the 45-day period. A day of face-to-face teaching lost is accrued for each student, for each day that they are unable to attend school, and is therefore independent of the structure of the school.

Sensitivity analyses

To enable the analysis to be applied across a wide range of contexts, sensitivity analyses were performed to examine how outcomes varied with different assumptions or inputs. The parameters that were varied were: school incursion rates (the model was initialized with one, two, or three simultaneous incursions as a proxy for community prevalence, where settings with high prevalence are more likely to have simultaneous or otherwise overlapping outbreaks due to high incursion rates from the community); and compliance with test-to-stay (also an equivalent sensitivity analysis for lower test sensitivity) ranging from 0 to 100%.

A number of other parameters were also varied, including vaccine coverage, non-pharmaceutical interventions, frequency of surveillance testing, number of non-classroom contacts, and symptomatic testing rate. These are provided in the Supplementary material.

Total days of face-to-face teaching gained

The modelled scenarios provide estimates of the days of face-to-face teaching lost per incursion. The total number of days lost or gained can be estimated based on the number of school incursions that take place. An example of how this calculation could be performed is provided in the Supplementary material.

Results

In this section, we simulate incursions in primary and secondary schools following an incursion event, and report on the number of downstream infections and days of face-to-face teaching lost under a range of control strategies. The outcomes described here are specific to the parameter values and assumptions outlined in the previous section, and therefore general trends should be considered for policy rather than the specific quantitative values.

Surveillance strategies, without contact tracing/quarantine

We first examined the impact of surveillance testing, in the absence of contact tracing. In many simulations, the initial incursion led to less than five downstream infections, and of these, the majority had no onwards transmission at all (Figure 2).

Twice weekly screening of teachers had minimal impact on reducing infections in primary schools, and only a marginal impact in

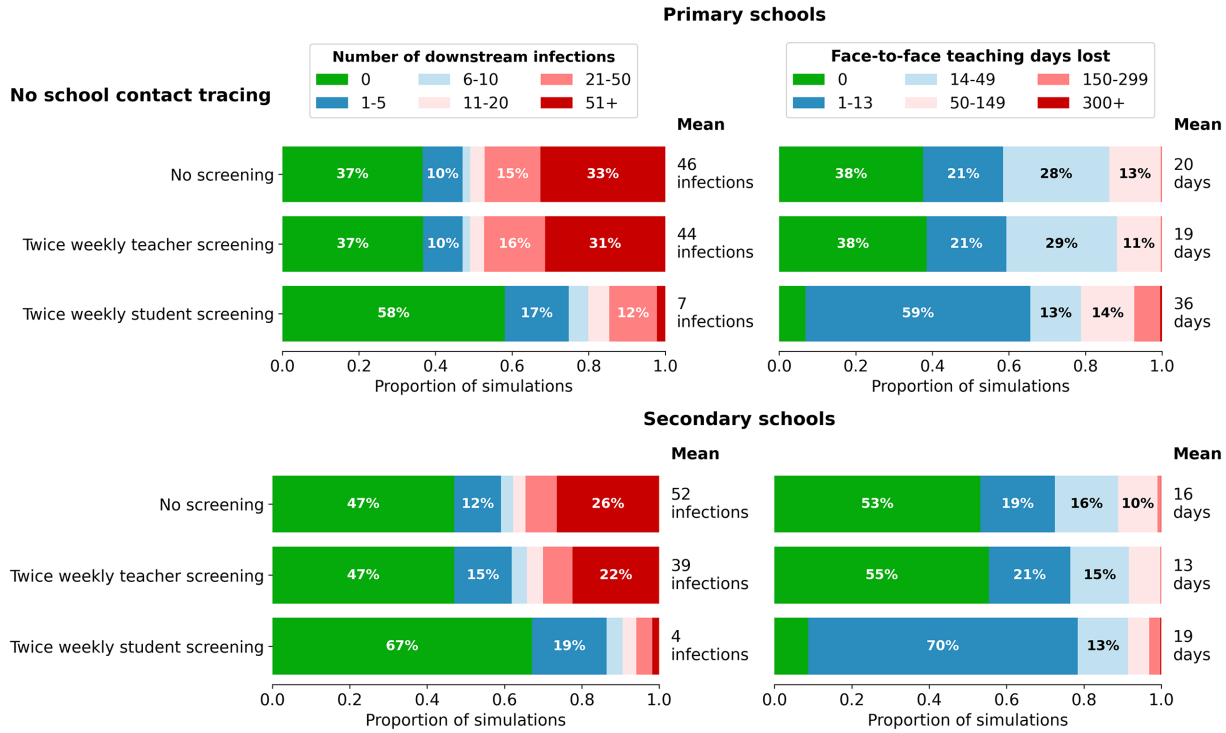


FIGURE 2
Impact of surveillance strategies on the distribution of outcomes for cumulative infections **(left)** and days of face-to-face teaching lost **(right)** in a single school following a single incursion. Outcomes are from 1,000 model simulations run for 45 days following first diagnosis. Scenarios assume no contact tracing or quarantine (only isolation for positive cases that are detected) and from top to bottom are based on: no screening; twice weekly testing of teachers with rapid antigen tests; twice weekly testing of students with rapid antigen tests.

secondary schools. Twice weekly screening of students led to earlier detection of an incursion and increased the chances of an incursion leading to no secondary infections in both primary and secondary schools. Screening of students slightly increased the *mean* days of face-to-face teaching lost compared with no screening and no contact tracing due to the detection of asymptomatic cases. However, the average masks the fact that this scenario resulted in a 20% increase in the proportion of incursions where transmission was effectively averted by early detection, a marked reduction in outbreaks of size 20 or more, and a reduction in the proportion of simulations with 14 or more days of face-to-face teaching lost.

Twice weekly screening of students had a greater impact on reducing secondary infections in schools as the number of incursions increased (Figure 3). With increased numbers of incursions, days of face-to-face teaching lost in secondary schools remained similar with or without student screening. In primary schools, days of face-to-face teaching lost slightly increased with the screening in place regardless of number of incursions. Overall, as the number of incursions increased, the incremental benefits for reduced secondary infections were far greater than the increase in days of face-to-face teaching lost.

Contact tracing and quarantine strategies: "test-to-stay"

We next examined contact tracing and the impact of testing contacts compared to quarantining contacts, in the absence of surveillance testing. Quarantining classroom contacts of identified cases considerably decreases the mean size of outbreaks after 45 days – from 46 cases to 26 cases in primary schools, and 52 cases to 25 cases in secondary schools (Figure 4). However, this comes at the expense of a large number of face-to-face teaching days lost per incursion – an average of 256 days in primary schools, and almost 700 days in secondary schools.

Test-to-stay of classroom contacts had approximately equivalent impacts on transmission as seven-day quarantine of classroom contacts in both primary and secondary schools, but without the associated face-to-face teaching days lost (Figure 4). Increased case ascertainment resulted in slightly more days of face-to-face teaching lost compared to the baseline no contact tracing scenario, but as with surveillance there was a marked reduction in the proportion of incursion leading to outbreaks of 20 or more. The effectiveness of test-to-stay decreased with low compliance, but conversely, there were diminishing returns at high levels of compliance (Figure 5).

The incremental benefit of testing all school contacts in addition to classroom contacts was small.

Surveillance strategies combined with contact tracing/quarantine

Finally, we examined the impact of surveillance testing, when combined with contact tracing and a test-to-stay strategy. Twice weekly screening of students had incremental benefits in terms of

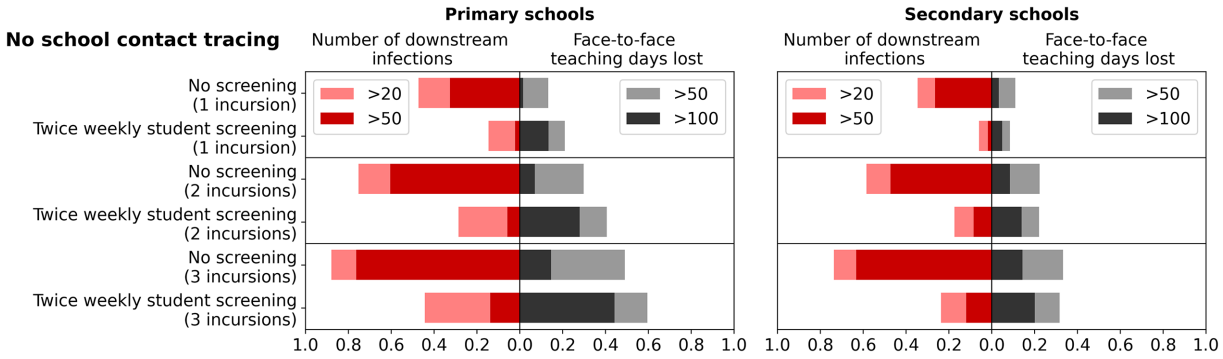


FIGURE 3
Impact of multiple incursions on the benefits of surveillance testing. Left bars: the percentage of simulations with more than 20 or 50 cumulative infections after 45days of first diagnosis, for different surveillance strategies and number of initial incursions. Right bars: the percentage of simulations with more than 50 or 100days of face-to-face teaching lost in a single school following the incursions. Outcomes are from 1,000 model simulations run for 45days following first diagnosis. Scenarios assume no contact tracing or quarantine (only isolation for positive cases that are detected) and compare no screening to twice weekly testing of students with rapid antigen tests.

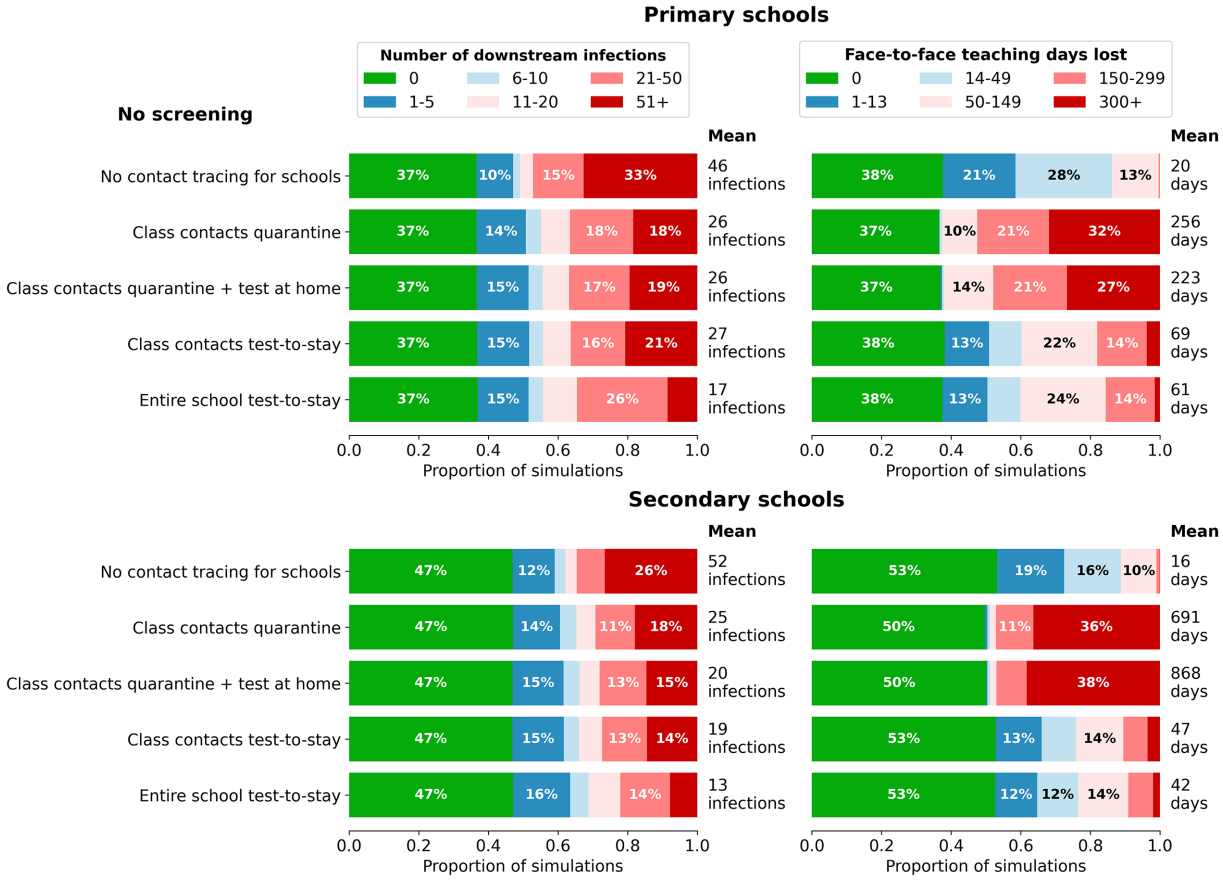


FIGURE 4
Impact of contact tracing and quarantine strategies on the distribution of outcomes for cumulative infections **(left)** and days of face-to-face teaching lost **(right)** in a single school following a single incursion. Outcomes are from 1,000 model simulations run for 45days following first diagnosis. Scenarios top to bottom: no contact tracing; class contacts have 7-day quarantine without/with testing; class contacts test-to-stay with rapid antigen tests; entire schools test-to-stay with rapid antigen testing. Top: Primary schools; bottom: Secondary schools.

reduced infections and reduced face-to-face teaching days lost (Supplementary material; Figure S1), even when test-to-stay was in place. By detecting and removing cases earlier, student screening reduces the number of downstream infections following an incursion, the likely outbreak size, and the average days of face-to-face teaching lost per incursion. In particular, screening resulted in a considerable reduction in the proportion of simulations where more than 150 days were lost.

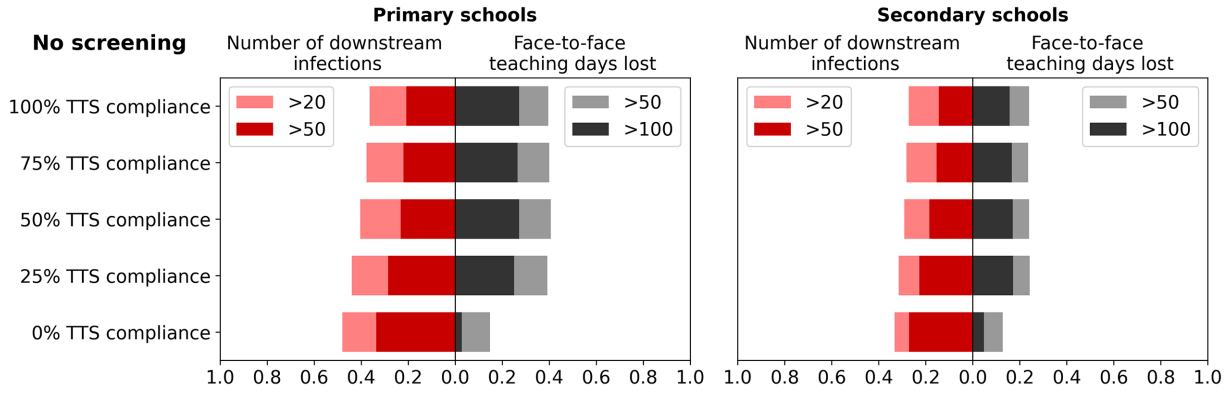


FIGURE 5
Impact of compliance on the effectiveness of a test-to-stay (TTS) strategy. Left bars: the percentage of simulations with more than 20 or 50 cumulative infections after 45days of first diagnosis, for different TTS compliance. Right bars: the percentage of simulations with more than 50 or 100 days of face-to-face teaching lost in a single school following the incursion. Outcomes are from 1,000 model simulations run for 45 days following first diagnosis.

TABLE 2 Model-estimated percentage change in mean number of infections and days of face-to-face teaching lost in primary and secondary schools compared to no intervention, for the scenarios shown in Figures 2, 4.

Intervention	Primary schools		Secondary schools	
	Mean infections	Days lost	Mean infections	Days lost
Screening interventions				
Twice weekly teacher screening	−4% (−12,0)	−5% (−12,1)	−25% (−34,-17)	−19% (−31,-12)
Twice weekly student screening	−85% (−87,-83)	80% (65,96)	−92% (−94,-91)	19% (2,38)
Tracing interventions				
Class contacts quarantine	−43% (−47,-42)	1180% (1119,1257)	−52% (−55,-50)	4219% (3932,4548)
Class contacts quarantine + test at home	−43% (−46,-40)	1015% (963,1086)	−62% (−65,-60)	5325% (4923,5774)
Class contacts test-to-stay	−41% (−45,-39)	245% (229,269)	−63% (−65,-60)	194% (170,221)
Entire school test-to-stay	−63% (−66,-61)	205% (188,223)	−75% (−77,-74)	163% (140,184)

No change | Moderate reduction | Large reduction | Moderate increase | Large increase

Results are presented as the percentage change relative to the no intervention scenario (95% confidence interval). Confidence intervals were calculated using the percentile bootstrap method with 1,000 replicates.

Summary of trends for surveillance and contact tracing strategies

The change in infections and days of face-to-face teaching lost for each of the control strategies and settings compared to no intervention is summarized in Table 2.

Discussion

This study used an agent-based model to assess the effectiveness of a variety of school-based surveillance, contact tracing and quarantine strategies to reduce outbreaks and transmission in schools, and maximize face-to-face teaching. We found that twice-weekly surveillance testing of students markedly reduced outbreak risk by enabling early detection of incursions, and that a 'test-to-stay' contact tracing strategy could achieve equivalent outbreak containment to home quarantine, without the associated loss of face-to-face teaching days. This was true for both primary and secondary schools.

School based surveillance testing considerably increased the proportion of simulations where an incursion resulted in no more than five downstream infections within 45 days. The incremental benefits of student surveillance testing were greater as the incursion rate increased, indicating that surveillance testing is expected to have maximum utility in areas with higher-than-average community transmission. Surveillance screening of students was also found to act synergistically with contact tracing and could be added on top of other policies such as test-to-stay to further reduce infections and in-person teaching days lost in areas at risk of outbreaks, but this would need to be balanced against the burden of testing for students and teachers.

The test-to-stay strategy outperformed a home quarantine strategy significantly in terms of maximizing face-to-face teaching. This is because a single infection in a class can result in more than 20 students losing 7 days each, and if transmission occurs to other classrooms (or in secondary school where students have multiple subjects) then quarantine requirements can be multiplicative. For test-to-stay to perform equivalently to home quarantine in terms of minimizing outbreak size, students must be compliant with testing requirements;

however, we found that much of the benefits were still realized with only 50% compliance. There are also concerns about the reduced sensitivity of rapid antigen tests to detect infections in classroom contacts, but in this appears to be compensated for with the high frequency of testing. The test-to-stay strategy was assessed with an assumed 80% vaccination coverage among students 12+ years, which was based on expected vaccine uptake at the time of analysis. The actual coverage of vaccines in Australia reached 77 and 88% among people aged 12–15 and 16–19 years by 31 January 2022 (21), suggesting this was a reasonable approximation, and sensitivity analyses showed that it would still be effective at lower vaccine coverage (as might be the case in international settings). Overall, the key finding that test-based strategies could provide epidemiological outcomes equivalent to quarantine was consistent with international empirical studies and evaluations that have occurred since (6, 7, 22, 23).

The report from the commissioned work was delivered to the Australian Government in November 2021 and was used to inform policy development and school-based strategies for 2022. The Australian school year aligns with the calendar year, and schools take summer vacations with most reopening in late January. In November–December 2021 the Omicron variant spread throughout Australia with cases peaking in early/mid-January 2022, substantially increasing prevalence in the community ahead of school reopening. In January 2022, state and federal governments developed the National Framework for Managing COVID-19 in Schools and Early Childhood Education and Care (24) affirming the importance of keeping schools open without prescribing specific policies as to how this was to happen. Accordingly, each state implemented their own return-to-school policies, depending on their specific health and education systems and local state of the epidemic. In January 2022, two states, South Australia and Western Australia, adopted the test-to-stay strategy, while Australia's two most populous states (NSW and Victoria) and those with the highest case incidence of SARS-CoV-2 implemented a twice-weekly screening program, and daily testing for 5 days for children in higher risk special school settings.

Strengths

A key strength of the modelling was that it was embedded as a part of the policy-making process. Modelers were engaged from the beginning, which enabled a deeper understanding of the questions most relevant to governments, and analyses could be collaboratively designed to best answer them. Another strength was that modelling was used to quantify key outcomes that could not be measured through data analysis or other methods. For example, quantifying the days of face-to-face teaching gained by early detection and isolation of infected students and the subsequent prevention of onward transmission. Finally, the modelling formed just one component of broader advice to government, who were therefore able to incorporate modelling outcomes alongside other forms of evidence, including data analysis, feasibility, acceptability, and logistic issues.

Limitations and future work

The findings presented are derived from an individual-based model, which is an imperfect representation of the real world with uncertainties in many parameters relating to disease progression and transmission. Model parameters were based on best-available data at the time of analysis.

The specificity of RATs has been measured in the range of 99.73–100% (17) and we therefore elected not to include false positive test results in this study. There would be additional days of face-to-face teaching lost due to false positive results associated with the surveillance testing and test-to-stay strategies. However, aside from the high test specificity limiting the number of false positives, we also note that many of the false positives arising from surveillance testing would occur in the absence of an active outbreak, and are therefore not captured within the scope of the simulations examined here.

Modelling was conducted based on the Delta variant. However, sensitivity analyses suggest that outbreak risks and days of face-to-face teaching lost following an incursion are greater with a more infectious variant. This makes early detection even more important with more infectious variants and means that the results of this study are likely to be even more pronounced than were estimated at the time. Reduced vaccine efficacy would be likely to increase the number of incursions and observed transmissibility, further accentuating this effect. The sensitivity analyses for TTS compliance (Figure 5) and screening frequency (Supplementary material) are equivalent to varying the test sensitivity, and suggest that a moderate reduction in RAT test sensitivity associated with new variants would be unlikely to qualitatively change our findings.

Specific to schools, limited network-type data on contact patterns within schools mean that mixing is approximated as consisting of classroom and non-classroom contacts, where students are allocated at random to classrooms and then randomly mix with other students outside of classrooms (rather than having social clustering). Some findings are also sensitive to assumptions for the number of non-classroom contacts students have; quarantine or test-to-stay strategies in particular focus on classroom contacts rather than close contacts as they are more practical to identify. However, these strategies will be less effective if a greater proportion of risk comes from non-class contacts.

Surveillance of teachers was found to have minimal benefit for reducing outbreaks in schools. Teachers only comprise a small proportion of the school community and for the purpose of this analysis we assumed that students and teachers had the same probability of becoming infected outside the school and causing the incursion. If teachers have a higher risk of becoming infected in the community than students, which may be the case (25), then the benefit of screening teachers would be higher than estimated.

Schools are embedded within their broader communities and receive incursions from the community as well as seeding cases back into the community. For this study, outbreaks were projected after a random initial incursion, without modelling the process by which the incursions occur. However, there may be social or other factors that make teachers or older/younger students more likely to be exposed in the community, and hence more likely to be the index case within the school, and these could change the effectiveness of different control measures. Limiting the scope of analysis to the outbreak within a school also meant that the benefits of community public health responses on reducing incursions into schools are not modelled, nor the benefits of school closures on reducing overall community transmission. Future work could assess the impact of community interventions on schools, and impact of school interventions on the rest of the community.

We have examined test-to-stay and quarantine protocols in the specific context of COVID-19 outbreaks, but the same questions are relevant for other respiratory infections such as influenza. The general principle of test-to-stay providing similar protection to quarantine is strongly dependent on test sensitivity, but is likely to also depend on disease attributes such as the incubation period, pre-symptomatic infectiousness, and whether there are asymptomatic carriers. Future work could investigate how such factors affect the impact of policy responses.

Conclusion

Twice-weekly surveillance testing of students with RATs can markedly reduce outbreak risk in schools by enabling early detection of incursions and is likely to have greatest benefit in areas with higher community transmission. Following an outbreak in a school, as an alternative to home quarantine a 'test-to-stay' strategy for class contacts achieves equivalent outbreak containment and enables face-to-face teaching. Evaluation of both approaches in schools will be critical to inform ongoing policy decisions and to optimize implementation of testing in educational settings when needed to reduce incursions.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

RA, JM, and NS conceived the study and developed the methodology. JM and NS conducted stakeholder consultations. RA, FR, MD, MH, JM, and NS devised scenarios. RA, RS-D, KH, DD, and NS developed the model. RA and NS conducted the analyses and drafted the manuscript. FR, MD, MH, and JM validated inputs and outputs. All authors contributed to the article and approved the submitted version.

Funding

RA, RS-D, KH, DD, and NS have received funding from the Victorian Department of Health (DoH), NSW DoH and the Federal Government for modelling related to COVID-19. FMR received funding from the Victorian Government's Department of Health and Human Services to analyze SARS-CoV-2 school outbreak data in 2020 and is a member of the Australia government's Department of Foreign Affairs and Trade Expert Advisory Group for Regional Vaccine Access and Health Security Initiative. MD received funding from the Victorian Government's Department of Health and Human Services to undertake enhanced public health investigation of SARS-CoV-2 cases in Victorian schools and early childhood education and care, was a member of the COVID-19 ATAGI working group on vaccine safety and confidence 2020–21 and is a member of the Australian Expert Technical Assistance Program for Regional COVID-19 Vaccine Access: Policy, Planning and Implementation (AETAP-PPI) Advisory Group. MH receives funding from the Victorian Government (DoH, Department of Families, Fairness and Housing, and Department of Jobs, Precincts and Regions) and the Macquarie Foundation to undertake work monitoring the impact of COVID-19 on the community. JM is an invited expert member of the Australian Health Protection Principal Committee, the Communicable Diseases Network of Australia COVID-19 working group and the Australian Technical Advisory Group on Immunization COVID-19 vaccine technical working group. This work was directly funded by the Australian Government Department of Health Office of Health Protection.

Acknowledgments

MH, JM, FR, and NS are the recipients of National Health and Medical Research Council fellowships. The authors gratefully acknowledge the support provided by the Victorian Government Operational Infrastructure Support Program.

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.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1150810/full#supplementary-material

References

1. Department of the Prime Minister and Cabinet (2021). Available at: https://www.pmc.gov.au/national-plan-transition-australias-national-covid-response (Accessed March 29, 2022).

2. Ryan K, Snow K, Danchin M, Mulholland K, Goldfeld S, Russell F. SARS-CoV-2 infections and public health responses in schools and early childhood education and care centres in Victoria, Australia: An observational study. *Lancet Regional Health-Western Pacific*. (2022) 19:100369. doi: 10.1016/j.lanwpc.2021.100369

3. Goldfeld S, O'Connor E, Sung V, Roberts G, Wake M, West S, et al. Potential indirect impacts of the COVID-19 pandemic on children: a narrative review using a community child health lens. *Med J Aust*. (2022) 216:364–72. doi: 10.5694/mja2.51368

4. Aiano F, Mensah AA, Mcowat K, Obi C, Vusirikala A, Powell AA, et al. COVID-19 outbreaks following full reopening of primary and secondary schools in England: Cross-sectional national surveillance, November 2020. *Lancet Regional Health-Europe*. (2021) 6:100120. doi: 10.1016/j.lanepe.2021.100120

5. Young BC, Eyre DW, Kendrick S, White C, Smith S, Beveridge G, et al. Daily testing for contacts of individuals with SARS-CoV-2 infection and attendance and SARS-CoV-2 transmission in English secondary schools and colleges: an open-label, cluster-randomised trial. *Lancet*. (2021) 398:1217–29. doi: 10.1016/S0140-6736(21)01908-5

6. Harris-McCoy K, Lee VC, Munna C, Kim AA. Evaluation of a Test to Stay Strategy in Transitional Kindergarten Through Grade 12 Schools—Los Angeles County, California, August 16–October 31, 2021. *Morb Mortal Wkly Rep*. (2021) 70:1773–7. doi: 10.15585/mmwr.mm705152e1

7. Nemoto N, Dhillon S, Fink S, Holman EJ, Cope AK, Dinh T-H, et al. Evaluation of Test to Stay Strategy on Secondary and Tertiary Transmission of SARS-CoV-2 in K–12 Schools—Lake County, Illinois, August 9–October 29, 2021. *Morb Mortal Wkly Rep*. (2021) 70:1778–81. doi: 10.15585/mmwr.mm705152e2

8. Kerr CC, Stuart RM, Mistry D, Abeysuriya RG, Rosenfeld K, Hart GR, et al. Covasim: An agent-based model of COVID-19 dynamics and interventions. *PLoS Comput Biol*. (2021) 17:e1009149. doi: 10.1371/journal.pcbi.1009149

9. Covasim model GitHub repository (2020). Available at: https://github.com/InstituteforDiseaseModeling/covasim.

10. Scott N, Palmer A, Delport D, Abeysuriya R, Stuart R, Kerr CC, et al. Modelling the impact of relaxing COVID-19 control measures during a period of low viral transmission. *Med J Aust*. (2021) 214:79–83. doi: 10.5694/mja2.50845

11. Abeysuriya RG, Delport D, Stuart RM, Sacks-Davis R, Kerr CC, Mistry D, et al. Preventing a cluster from becoming a new wave in settings with zero community COVID-19 cases. *BMC Infect Dis*. (2022) 22:1–15. doi: 10.1186/s12879-022-07180-1

12. Abeysuriya R, Delport D, Sacks-Davis R, Hellard M, Scott N. (2021). Modelling the Victorian roadmap. Public Health Report. Available at: https://www.burnet.edu.au/media/l23ekpol/burnet_institute_vic_roadmap_20210918_-final.pdf (Accessed September 18, 2021).

13. Australian Bureau of Statistics (ABS) (2020). Education statistics. Available at: https://www.abs.gov.au/statistics/people/education/schools/2020 (Last accessed May 18, 2023).

14. Victorian Government (2021). Government schools average class size. Available at: https://discover.data.vic.gov.au/dataset/government-schools-average-class-size-primary-classes-feb-victoria (Last accessed May 18, 2023).

15. OECD (2021). *Education at a Glance 2021: OECD Indicators*, OECD Publishing, Paris. doi: 10.1787/b35a14e5-en

16. Davies NG, Klepac P, Liu Y, Prem K, Jit M, Pearson CAB, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat Med*. (2020) 26:1205–11. doi: 10.1038/s41591-020-0962-9

17. Muhi S, Tayler N, Hoang T, Ballard SA, Graham M, Rojek A, et al. Multi-site assessment of rapid, point-of-care antigen testing for the diagnosis of SARS-CoV-2 infection in a low-prevalence setting: A validation and implementation study. *Lancet Regional Health-Western Pacific*. (2021) 9:100115. doi: 10.1016/j.lanwpc.2021.100115

18. Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, Zambrano-Achig P, Del Campo R, Ciapponi A, et al. False-negative results of initial RT-PCR assays for COVID-19: a systematic review. *PLoS One*. (2020) 15:e0242958. doi: 10.1371/journal.pone.0242958

19. Pouwels KB, Pritchard E, Matthews PC, Stoesser N, Eyre DW, Vihta K-D, et al. Effect of Delta variant on viral burden and vaccine effectiveness against new SARS-CoV-2 infections in the UK. *Nat Med*. (2021) 27:2127–35. doi: 10.1038/s41591-021-01548-7

20. Eyre DW, Taylor D, Purver M, Chapman D, Fowler T, Pouwels KB, et al. Effect of Covid-19 Vaccination on Transmission of Alpha and Delta Variants. *N Engl J Med*. (2022) 386:744–56. doi: 10.1056/NEJMoa2116597

21. Department of Health (2022). COVID-19 vaccine rollout update – 31. Available at: https://www.health.gov.au/sites/default/files/documents/2022/01/covid-19-vaccine-rollout-update-31-january-2022.pdf (Accessed April 08, 2022).

22. Young BC, Eyre DW, Kendrick S, White C, Smith S, Beveridge G, et al. Daily testing for contacts of individuals with SARS-CoV-2 infection and attendance and SARS-CoV-2 transmission in English secondary schools and colleges: an open-label, cluster-randomised trial. *Lancet*. (2021) 398:1217–29. doi: 10.1016/S0140-6736(21)01908-5

23. Campbell MM, Benjamin DK, Mann T, Fist A, Kim H, Edwards L, et al. Test-to-Stay After Exposure to SARS-CoV-2 in K–12 Schools. *Pediatrics*. (2022) 149:e2021056045. doi: 10.1542/peds.2021-056045

24. National Framework for Managing COVID-19 in Schools and Early Childhood Education and Care (2022). Department of Education, Skills and Employment. Available at: https://www.education.gov.au/covid-19/resources/national-framework-managing-covid19 (Last accessed May 18, 2023).

25. National Centre for Immunisation Research and Surveillance (2022). COVID-19 in schools – the experience in NSW: 18 October 2021 to 17 December 2021. February 2022 Report. Available at: https://www.ncirs.org.au/covid-19-in-schools (Accessed May 6, 2022).

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 05 July 2023
DOI 10.3389/fpubh.2023.1193100

OPEN ACCESS

EDITED BY
Pierpaolo Ferrante,
National Institute for Insurance against Accidents at Work (INAIL), Italy

REVIEWED BY
Pei Yuan,
York University, Canada
Syed Muzamil Basha,
REVA University, India

*CORRESPONDENCE
David Soriano-Paños
✉ sorianopanos@gmail.com
Jesús Gómez-Gardeñes
✉ gardenes@unizar.es

RECEIVED 24 March 2023
ACCEPTED 19 June 2023
PUBLISHED 05 July 2023

CITATION
Valgañón P, Lería U, Soriano-Paños D and Gómez-Gardeñes J (2023) Socioeconomic determinants of stay-at-home policies during the first COVID-19 wave.
Front. Public Health 11:1193100.
doi: 10.3389/fpubh.2023.1193100

Socioeconomic determinants of stay-at-home policies during the first COVID-19 wave

Pablo Valgañón[1,2], Unai Lería[1], David Soriano-Paños[2,3]* and Jesús Gómez-Gardeñes[1,2]*

[1]Department of Condensed Matter Physics, University of Zaragoza, Zaragoza, Spain, [2]GOTHAM Lab - Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Zaragoza, Spain, [3]Institute Gulbenkian of Science (IGC), Oeiras, Portugal

Introduction: The COVID-19 pandemic has had a significant impact on public health and social systems worldwide. This study aims to evaluate the efficacy of various policies and restrictions implemented by different countries to control the spread of the virus.

Methods: To achieve this objective, a compartmental model is used to quantify the "social permeability" of a population, which reflects the inability of individuals to remain in confinement and continue social mixing allowing the spread of the virus. The model is calibrated to fit and recreate the dynamics of the epidemic spreading of 42 countries, mainly taking into account reported deaths and mobility across the populations.

Results: The results indicate that low-income countries have a harder time slowing the advance of the pandemic, even if the virus did not initially propagate as fast as in wealthier countries, showing the disparities between countries in their ability to mitigate the spread of the disease and its impact on vulnerable populations.

Discussion: This research contributes to a better understanding of the socioeconomic and environmental factors that affect the spread of the virus and the need for equitable policy measures to address the disparities in the global response to the pandemic.

KEYWORDS

COVID-19, epidemic modeling, Bayesian inference, compartmental models, non-pharmaceutical containment policies

1. Introduction

The COVID-19 disease, caused by the novel coronavirus SARS-CoV-2, is a highly contagious respiratory illness that was first reported in Wuhan, China in December 2019 (1). It was declared a pandemic by the World Health Organization on March 11th 2020 and by that time, the disease had spread globally, resulting in an international public health crisis that impacted all aspects of life for those affected. As it continued to spread, countries worldwide implemented a range of strict measures to contain the virus with varying degrees of success, including lockdowns, travel restrictions, and social distancing measures (2). Since the early stages of the pandemic, experts have studied the uneven spatial spread of the virus, which is associated with socioeconomic and environmental factors (3, 4). These studies revealed that minorities, low-income areas, and vulnerable populations (5, 6) have been disproportionately affected by the situation, exacerbating existing inequalities.

Numerous studies have examined the factors that contribute to the reproduction number or the speed at which a disease propagates in a population (7). Most of these studies have found positive correlations between the transmissibility of a pathogen and factors such as population density, income inequality, urban areas, and household size, among others (8–11). At the onset of the COVID-19 pandemic, an important metric that also

showed a positive correlation with the reproduction number of SARS-CoV-2 was the Gross Domestic Product (GDP) per capita (12–14). This suggests that the virus spread more rapidly in affluent countries due to a large percentage of the population living in densely populated cities and the significant inflow of air traffic facilitating the initial importation of a large number of cases. However, the data on this correlation was inconclusive in the subsequent phases of the COVID-19 pandemic.

Despite the implementation of measures to mitigate the spread of the virus, there are significant disparities between countries in their ability to contain the pandemic. While wealthier countries have been able to enforce stay-at-home policies by taking the appropriate measures and ensuring the safety of the most vulnerable populations, lower income countries are unable to prevent the loss of income and jobs, leading to food insecurity and reduced access to healthcare. As a result, even with mobility restrictions in place, people in these countries are still exposed to the risk of the pandemic due to the need to work and maintain their income.

The goal of this study is to highlight the connection between the different efficacy of lockdown policies observed across countries and their socioeconomic features. To this aim, we will attempt to measure the *social permeability* of these nations, which accounts for the inability of the population to remain in confinement and thus continuing social mixing that allows the disease to spread. The study of this variable allow us to distinguish unique scenarios that appear for each studied country and contribute to the already existing efforts (7) to show the varied relationships between the spread of epidemics and economic indicators.

2. Methods

2.1. Modeling

2.1.1. Discrete time compartmental model

The main core of this study is the development of a compartmental model to capture COVID-19 epidemic trajectories and how they were impacted by non-pharmaceutical interventions implemented in various countries. To this aim, the model should be sufficiently complex to provide an accurate representation of the epidemic process and the primary mechanisms behind virus transmission, yet simple and adaptable enough to be applied to different countries.

The proposed compartmental framework is an extended version of the *Susceptible-Exposed-Infected-Recovered* (SEIR) model (15), which allows for monitoring both the number of deaths over time for each country and the effects of non-pharmaceutical interventions. Our framework is a discrete-time model, being each time step a day. The model consists of six compartments: *Susceptible* (S), *Exposed* (E), *Infectious* (I), *Recovered* (R), *Pre-deceased* (P_d), and *Deceased* (D). The flows diagram connecting these compartments is represented in Figure 1. This diagram depicts the following sequence of events: In the absence of interventions, *Susceptible* individuals (S) have a likelihood of contracting the virus (β) for each contact with an infected person, leading them to move to the *Exposed* compartment (E), meaning that they are carriers of the virus but not yet contagious. Once in compartment E, individuals can move to the *Infectious* compartment (I) with a probability η, where they can infect *Susceptible* individuals. *Infectious* individuals leave their compartment with probability μ, either recovering and entering the *Recovered* compartment (R) with a probability of $1-\Upsilon$, or dying due to the disease with a probability of Υ. In the latter case, they enter in the *Pre-deceased* compartment (P_d) and, eventually, move to the *Deceased* compartment (D) with probability ξ.

Following the previous assumptions for the compartmental model, we can propose the dynamical equations driving the evolution of the individuals. In particular, the evolution from the *Infectious* stage can be straightforwardly derived, yielding:

$$I(t+1) = \eta E(t) + (1-\mu)I(t)\,, \tag{1}$$
$$R(t+1) = \mu(1-\Upsilon)I(t) + R(t)\,, \tag{2}$$
$$P_d(t+1) = \mu\Upsilon I(t) + (1-\xi)P_d(t)\,, \tag{3}$$
$$D(t+1) = \xi P_d(t) + D(t)\,. \tag{4}$$

The remaining equations of the model account for the contagion of the *Susceptible* individuals and are shaped by the non-pharmaceutical interventions, which manifest in a reduction in mobility and the formation of *social bubbles* throughout the populations. The impact of non-pharmaceutical interventions will be encapsulated in a fraction of the *Susceptible* population gathering the number of individuals staying at their households, not being reachable by their infectious counterparts. To model the evolution of this confined population, we assume that mobility is governed by a time-dependent parameter $p_{act}(t)$. In particular, for each day, a fraction $p_{act}(t)$ of the *Susceptible* compartment remains active (S_{active}) while the rest, a fraction $(1-p_{act}(t))$ of the pool of *Susceptibles*, reduce their mobility and social interactions. Of these individuals, a fraction $(1-\phi)$ stays at home and forms a social bubble with the rest of the members of the household ($S_{confined}$). The rest of the individuals that became inactive but did not successfully isolate themselves completely ($S_{inactive}$) mix with members of other households due to their *social permeability*. These three fractions of the *Susceptible* compartment can be represented as:

$$S_{active}(t) = S(t)p_{act}(t) \tag{5}$$
$$S_{inactive}(t) = S(t)(1-p_{act}(t))\phi \tag{6}$$
$$S_{confined}(t) = S(t)(1-p_{act}(t))(1-\phi) \tag{7}$$

Taking into account the aforementioned policies and every group in the *Susceptible* population, the equation governing the time evolution of the occupation of the *Exposed* compartment reads:

$$E(t+1) = S_{active}(t)P_{active}(t) + S_{inactive}(t)P_{inactive}(t) + S_{confined}(t)P_{confined}(t) + (1-\eta)E(t)\,, \tag{8}$$

where $P_{active}(t)$, $P_{inactive}(t)$ and $P_{confined}(t)$ account for the probability that *Susceptible* individuals belonging to each of these groups contract the disease at time t. We assume that the number of contacts made by each group ($k_{active}(t)$ and $k_{inactive}(t)$) depends on the mobility levels in their respective settings, yielding:

$$k_{active}(t) = \langle k_{active}\rangle p_{act}(t)\,, \tag{9}$$
$$k_{inactive}(t) = \langle k_{inactive}\rangle p_{res}(t)\,, \tag{10}$$

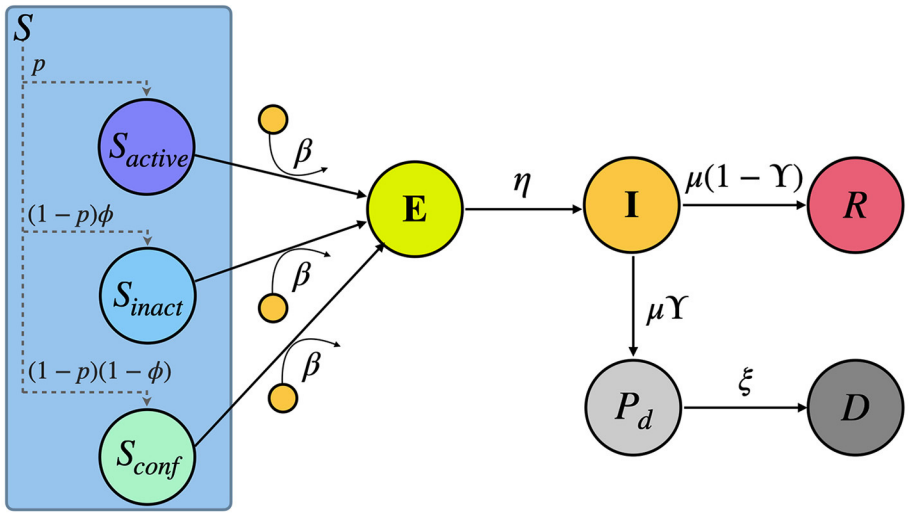


FIGURE 1
Scheme of the compartmental model here proposed. The model comprises six compartments: *Susceptible S*, *Exposed E*, *Infectious I*, *Recovered R*, *Pre-deceased* P_d and *Deceased D*. Note that, as a result of the non-pharmaceutical interventions, the *Susceptible* compartment is divided into three sub-compartments: S_{active}, $S_{inactive}$, and $S_{confined}$ representing a fraction $p(t)$, $(1-p(t))\phi$ and $(1-p(t))(1-\phi)$ of the total number of *Susceptible* individuals, respectively. A detailed explanation of the flows connecting these compartments can be found in the Section 2.

where $p_{act}(t)$ represents the observed mobility of the people that travel to their daily destinations and $p_{res}(t)$ is the mobility of those who remain inactive in their residential areas. Likewise, $\langle k_{active} \rangle$ corresponds to all contacts made by individuals in the baseline scenario whereas $\langle k_{inactive} \rangle$ constitutes their interactions at home. Both can be estimated from social data existing in the literature (16).

Assuming a well-mixed population, the probabilities of contracting the disease can be calculated as:

$$P_{active}(t) = 1 - \left(1 - \beta \frac{I(t)}{N}\right)^{k_{active}(t)}, \quad (11)$$

$$P_{inactive}(t) = 1 - \left(1 - \beta \frac{I(t)}{N}\right)^{k_{inactive}(t)}, \quad (12)$$

for the active and inactive population. In addition, there is a chance that confined individuals contract the disease from other infectious members of their social bubble. These agents make $k_{inactive}$ contacts with others in their own household. The probability of getting infected for this group of individuals depends on the number of infected people in their households as:

$$P_{confined}(t) = 1 - \sum_{i=0}^{\sigma-1} p(i) \left(1 - \beta \frac{i}{\sigma - 1}\right)^{k_{inactive}(t)}, \quad (13)$$

where σ is the number of people in each household (meaning that a susceptible member is able to make contacts with the other $\sigma - 1$ residents) and the probability of finding i infected individuals in a household is

$$p(i) = \binom{\sigma - 1}{i} \left(\frac{I(t)}{N}\right)^{i} \left(1 - \frac{I(t)}{N}\right)^{\sigma - 1 - i}. \quad (14)$$

To round off, we assume a closed population so that the occupation of the S compartment changes with:

$$S(t+1) = N - E(t+1) - I(t+1) - R(t+1) - P_d(t+1) - D(t+1). \quad (15)$$

2.2. Data sources

2.2.1. COVID-19 deaths

To calibrate our model, we rely on data regarding the daily number of fatalities in each country. As consistency across populations is crucial, the number of detected cases is not a suitable metric due to surveillance issues that affect countries differently (17, 18). There are several contributing factors to the possible discrepancies on the reported cases. Firstly, testing strategies can vary across populations, with some focusing on high-risk groups or areas, while not accounting for asymptomatic cases. Additionally, limited testing capacity can lead to an underestimation of the true number of infections, particularly in areas with high community transmission. Furthermore, differences in the definition and reporting of cases, as well as demographic variations such as age, gender, and underlying health conditions, can make it challenging to compare trends between different countries or regions.

The accuracy of reported deaths is also not guaranteed for similar reasons, including possible changes in definition and reporting delays. To address these reporting issues, we specifically choose countries with continuous and consistent records of this information, which are listed in Table 1. The selection of the countries was performed according to two criteria: there must be a peak of at least 10 deaths, and the daily mobility and number of deaths have to be available and consistent for this entire duration.

10.3389/fpubh.2023.1193100

TABLE 1 Average contacts of active, $\langle k_{active} \rangle$, and inactive, $\langle k_{inactive} \rangle$, average household size σ, GDP per capita and minimum level of mobility during lockdown for each country.

Country	Country code	$\langle k_{active} \rangle$	$\langle k_{inactive} \rangle$	σ	GDP per capita (US$)	p_{min}
Argentina	AR	14.12	3.84	2.95	8475.73	0.15
Austria	AT	12.48	3.13	2.27	48105.63	0.17
Bangladesh	BD	16.34	3.55	4.26	2000.64	0.25
Belgium	BE	11.38	2.88	2.36	45028.32	0.21
Bolivia (Plurinational State of)	BO	17.04	3.03	3.53	3133.1	0.12
Bulgaria	BG	12.73	4.06	2.34	10058.08	0.35
Canada	CA	12.57	3.1	2.45	43559.71	0.43
Chile	CL	13.71	3.66	3.04	13231.71	0.31
Colombia	CO	15.26	3.6	3.53	5332.77	0.18
Egypt	EG	15.67	3.64	4.13	3608.84	0.38
France	FR	11.78	3.1	2.22	38958.6	0.12
Germany	DE	6.86	1.79	2.05	45908.72	0.38
Greece	GR	11.79	3.18	2.44	18117.07	0.21
Guatemala	GT	18.98	4.08	4.81	4331.69	0.31
Honduras	HN	18.27	4.23	3.87	2405.73	0.17
Hungary	HU	12.07	3.46	2.6	16128.65	0.39
Indonesia	ID	15.26	3.14	3.86	3869.59	0.54
Iraq	IQ	20.64	4.35	6.35	4145.86	0.33
Ireland	IE	12.47	3.43	2.83	86250.99	0.15
Israel	IL	13.6	3.84	3.14	47033.59	0.14
Italy	IT	14.37	2.94	2.4	31238.05	0.11
Kuwait	KW	16.42	4.16	5.8	24809.04	0.12
Luxembourg	LU	16.37	3.46	2.41	117181.7	0.17
Malaysia	MY	15.4	3.59	4.56	10401.79	0.2
Mexico	MX	15.42	4.04	3.75	8325.57	0.4
Morocco	MA	14.3	3.72	4.58	3108.18	0.18
Nigeria	NG	20.47	4.16	4.66	2085.47	0.44
Pakistan	PK	18.65	4.18	6.8	1167.22	0.31
Panama	PA	14.62	3.45	3.64	12269.05	0.14
Philippines	PH	17.04	3.72	4.23	3298.83	0.17
Poland	PL	13.93	4.42	2.81	15764.11	0.32
Portugal	PT	11.88	3.03	2.66	22413.04	0.23
Romania	RO	11.97	3.41	2.88	12928.58	0.27
Russian Federation	RU	12.88	3.42	2.58	10165.51	0.45
Saudi Arabia	SA	15.64	4.1	5.6	20110.32	0.27
South Africa	ZA	15.92	3.94	3.36	5094.38	0.24
Spain	ES	12.02	3.19	2.69	27408.63	0.08
Switzerland	CH	13.14	3.11	2.21	86918.65	0.19
Turkiye	TR	13.72	3.72	4.07	8538.13	0.27
Ukraine	UA	12.65	3.48	2.53	3557.48	0.47
The United Kingdom	GB	9.48	2.25	2.27	40718.22	0.22
United States of America	US	12.6	3.24	2.49	63122.59	0.54

The data used in this study is extracted from the official daily counts of COVID-19 deaths reported for countries by the World Health Organization and smoothed using a 7-day rolling average.

2.2.2. Mobility reduction data

We extract the level of mobility inside each country at a certain time, $p_{act}(t)$ and $p_{res}(t)$, from the Google COVID-19 Community Mobility Reports. Among the different types of movements included in this study, we focus on the *Retail and Recreation* category for the active individuals and the *Residential* category for the inactive ones. For each day, the level of mobility is computed by comparing the amount of flows recorded for this day with their median values measured in a pre-pandemic baseline scenario, spanning 5 weeks from January 3 to February 6, 2020. To reduce data noise, we smooth the curves using a 7-day rolling average.

2.2.3. Socioeconomic data

A key component of the epidemiological model is the average number of people an individual encounters (contacts) throughout their day, which varies from one country to another and has been measured in a certain number of them taking into account the heterogeneities in the populations. The dataset used for the simulation comes from a study (16) where the authors extrapolate the known data to 152 countries and provide contact matrices representing the number of contacts a person of each age group has with the others in different settings. For each country, we obtain the average number of contacts from an active individual $\langle k_{active}\rangle$ as the weighted average of the total number of contacts made by each individual of each group in all the settings, taking into account the population age pyramid of this country (19). To compute the same quantity for inactive (controlled) individuals $\langle k_{inactive}\rangle$, we repeat the same process by just accounting for the contacts made at home.

The average number of residents in a single home is also an important parameter of the model, and was reported by the United Nations (20) for most countries in the world. The household size σ is available at https://www.un.org/development/desa/pd/data/household-size-and-composition, and in the model has been rounded to the nearest integer in order to follow the equations.

Lastly, the Gross Domestic Product (GDP) per capita is taking into account to find a relation between the wealth of different populations and the success of their confinement policies. It is available at www.worldbank.org.

These country-dependent parameters ($\langle k_{active}\rangle$, $\langle k_{inactive}\rangle$ and the GDP per capita) are summarized in Table 1.

2.2.4. Epidemiological parameters

Some of the parameters in relation to the compartmental model have already been determined and are fixed based on the literature:

- η: Probability of leaving the E compartment. It is related to the average duration of the incubation period. We fix its value to $\eta = 1.0/5.2$ (21).
- μ: Probability of leaving the I compartment. It is related to the average duration of the infectious windows after contracting and incubating the virus. We fix its value to $\mu = 1.0/4.2$ (21).
- Υ: Infection fatality rate which, as reported in (22) and (23), is estimated to be $\Upsilon = 0.01$.

2.3. Model calibration

2.3.1. Approximate Bayesian Computation (ABC)

The Approximate Bayesian Computation (ABC) method, as described in (24) and (25), provides a solution to Bayesian inference problems where computing the likelihood function and its further exploration becomes cumbersome. ABC works by generating synthetic trajectories using a set of parameters and then accepting or rejecting those parameters based on how well the synthetic trajectories match real data. This approach allows the construction of approximate posterior distributions.

There are several ways of exploring the posterior distribution of the parameters, one of the simplest being the ABC rejection algorithm (26), which is used in our case. To quantify the goodness of a given trajectory generated by a set of parameters $\vec{\theta}$, we use a logarithmic distance function $\rho(\vec{\theta})$ defined as:

$$\rho(\vec{\theta}) \equiv \sum_t \log\left[|D_{obs}(t) - D_{\vec{\theta}}(t)| + 1\right] , \tag{16}$$

where $D_{obs}(t)$ represents the observed daily fatalities at time t and $D_{\vec{\theta}}(t)$ its value predicted by the synthetic trajectory. Note that, among all possible choices for this goodness function, we have chosen a logarithmic function not to under-represent the initial stage of the epidemics, where there are fewer deaths.

The ABC rejection algorithm builds the posterior distribution for the model parameters by sampling them from the trajectories fulfilling that $\rho(\vec{\theta}) < \epsilon$, where ϵ is a tolerance threshold. In our case, we run two rounds of the ABC rejection algorithm. In the first round, we draw $20 \cdot 10^6$ random samples of variables from the prior distributions and compute the distance between the synthetic trajectories and the real data by taking into account the period between February 20 and May 20, 2020. We set a dynamical threshold ϵ_i to accept those 1,000 trajectories providing the best fits for the data in each country. We construct the prior distributions for the second from the accepted trajectories in the first one and the process is repeated. This second iteration allows the algorithm to give more accurate results for each country due to the intrinsic variability of the parameters across countries.

2.3.2. Model free parameters

The numerical iteration of Equation (1) allows one to obtain synthetic trajectories capturing the evolution of individuals in each of the compartments. The parameters of the model that are not fixed from the literature will be left for calibration via the ABC method. These parameters are:

- β: This parameter represents the probability of infection, which varies from one country to another due to factors such as population density, urbanization, and use of masks. The prior distribution of this parameter is $\beta \sim U(0.01, 0.3)$.
- ϕ: The permeability of the confinement, which is the main objective of the study to fit. A low permeability means a high

effectiveness of the non-pharmaceutical policies due to a good compliance from the population. The prior distribution of this parameter is $\phi \sim U(0, 1)$.

- ξ: The probability of leaving the P_d compartment to die because of the disease. The prior distribution of this parameter is $\xi \sim U(1/18, 1/6)$.
- T: The estimated number of days that have elapsed since the first case of the disease in the country and the day chosen as the starting point for comparison with observed deaths, which for all countries corresponds to 2020-02-20. The prior distribution of this parameter reads $T \sim U(0, 100)$.
- δ: The delay in death notification. The prior distribution of this parameter is set at $\delta \sim U(2, 20)$.

Note that δ is not a strictly needed parameter to run the model but becomes essential to make synthetic trajectories compatible with real data (27). In all the cases, the prior distributions chosen are broad enough to avoid biasing the inference of the posterior distributions.

2.4. Relationship between GDP per capita and countries permeability

In this section, we explain how we link the inferred permeability distributions of individual countries with certain economic indicators, such as GDP per capita. To obtain a meaningful association, we should exclude countries for which the model does not provide a reasonable fit, as well as those where mobility limitations did not substantially impact the control of the epidemic.

On the one hand, to determine which countries have been correctly modeled, we calculate the relative area between the estimated curve and the real data using the following equation:

$$\varepsilon(\vec{\theta}) = \frac{\sum_t |D_{obs}(t) - D_{\vec{\theta}}(t)|}{\sum_t D_{obs}(t)} \,. \tag{17}$$

For the subsequent analysis, we discard those countries for which $\min(\varepsilon(\vec{\theta})) > 0.4$ as the model does not fit well the epidemic trajectories there.

Once these countries are discarded, we quantify the relationship between permeability and GDP per capita by performing a non-linear regression fitting the permeability to the following function:

$$\phi(x) = ax^{-b} \tag{18}$$

where x is the GDP per capita. As our information about permeability comes from posterior distributions, we conduct 1, 000 independent fits by sampling diverse sets of permeability values from these distributions. The confidence interval of the regression curve is calculated as the percentile 2.5 to the percentile 97.5 of the individuals fits obtained.

3. Results

We calibrate our model to real data using the Approximate Bayesian Computation (ABC) scheme described in the Methods section. The results of the calibration for each considered country can be found in Supplementary Figure 1. One important finding is that, despite its simplicity, our simple model accurately captures the time evolution of reported deaths for most countries and confirms the assumption that the mobility reduction has a direct effect on the number of daily contagions, which decreases as stricter policies are put in place. Figure 2 illustrates this by showing the real and simulated epidemic trajectories for Spain, Colombia, and Ukraine. The selected countries represent three distinct types of behavior observed in our study. While Spain and Colombia implemented similar lockdown policies resulting in comparable reductions in average mobility, their outcomes were vastly different: Spain managed to stop the spread and bend the curve, whereas Colombia experienced a steady growth in casualties. This pattern of steady increase is also observed in Ukraine, which had a milder reduction in mobility compared to Spain and Colombia.

The unequal impact of mobility reduction on epidemic containment can be captured by the social permeability ϕ parameter in our model, which modulates the effective reproductive number of the circulating virus, as explained in (28). Namely, low permeability values significantly reduce the pool of susceptible individuals exposed to the virus due to the lower household mixing, whereas high permeability values means that all the individuals remain vulnerable to the virus but with a reduced exposure due to their hampered social activity.

In Figure 3, we present the posterior distributions obtained of social permeability for each country analyzed in this study. Focusing on the specific case of Colombia and Spain, we confirm that lower efficiency of mobility reductions in Colombia translates to higher permeability values compared to those inferred for Spain. While not the focus of our manuscript, other model parameters also provide insightful information about the impact of the first COVID-19 epidemic wave and the associated contention measures across countries. For instance, the inferred values of parameter T enable to reconstruct the time of onset of the outbreak in each country, whereas parameter δ accounts for the heterogeneous delay in reporting deaths. Nonetheless, conclusions on these parameters should be drawn with caution because of the correlations between their posterior distributions, as illustrated in Supplementary Figures 2–4 for the case of Spain, Colombia and Ukraine, respectively.

To round off, we check whether we can connect the heterogeneous permeability values inferred for each country with their corresponding socioeconomic features. In order to establish a meaningful link, we narrow our focus to countries where the model accurately predicts the course of the disease. We determine this accuracy by calculating the normalized distance $\varepsilon(\vec{\theta})$ between the data and the model trajectories, which enables us to establish a threshold and exclude countries where the model does not perform well. This procedure is described in more detail in the Section 2 and the distribution of the minimum normalized distances $\varepsilon^{\min}$ observed across countries is represented in Supplementary Figure 5.

Figure 4 represents the posterior distribution of the social permeability against the gross domestic product (GDP) per capita of the selected countries. The tendency showcased in the figure indicates that there is in fact a negative statistically significant correlation between the wealth of a country and the ability of its inhabitants to properly follow the restrictions and stay in lockdown.

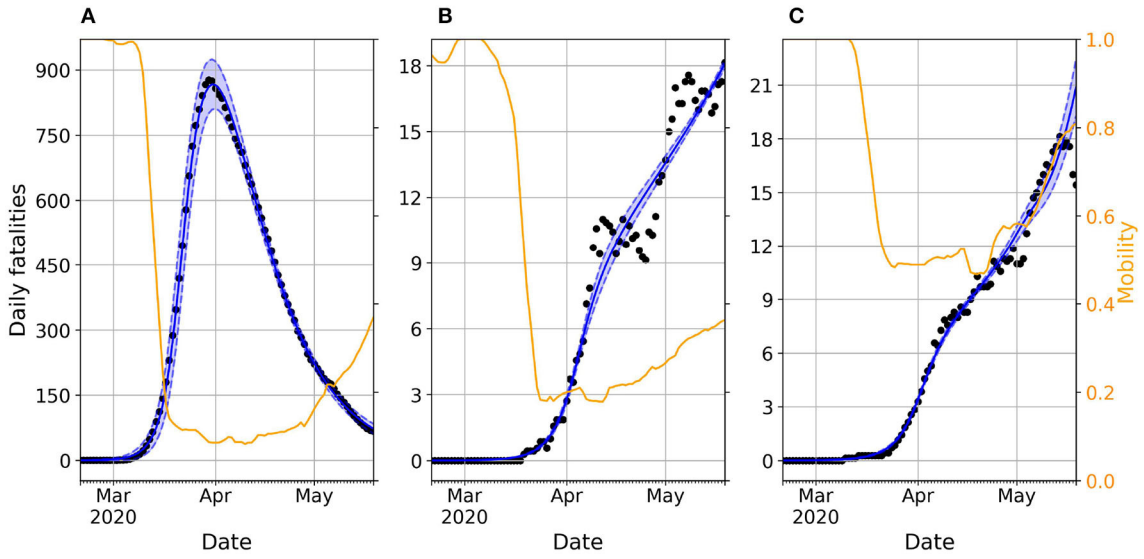


FIGURE 2
Daily evolution of the number of deaths in Spain **(A)**, Colombia **(B)**, and Ukraine **(C)**. In all the panels, dots represent real reported data whereas the blue shadowed region corresponds to the 95% prediction interval of the accepted trajectories after calibrating the model. The blue solid line represents the median trajectory whereas the orange line corresponds to the time variation of mobility compared with a baseline pre-pandemic scenario spanning from January 3 to February 6, 2020.

This negative correlation between the permeability and GDP per capita is further supported by the non-linear regression of the data described in the Section 2.

For the sake of completeness, we study the influence of possible confounding factors on this correlation such as mobility reduction and deaths caused by the disease. Supplementary Figure 6 shows that the permeability values have no correlation with the minimum observed mobility for each one of them, meaning the model can separate the level of mobility reduction and the effectiveness of the confinement without one depending on the other. Regarding the relationship between permeability and death toll per capita in each country, we observe in Supplementary Figure 7 low permeability values for those countries with higher number of fatalities but a large variability without any clear relationship in those less severely affected.

4. Discussion

The COVID-19 pandemic has had an undeniable impact on the world and has exposed the existing social and economic inequalities within many countries (29, 30). To address this issue, various countries have implemented policies and non-pharmaceutical interventions to control the virus's spread and reduce the number of casualties. However, the pandemic's impact has not been evenly distributed across society, with certain groups suffering more severe consequences than others (31, 32). This inequality is not limited to individual countries, but also occurs across nations due to various challenges that low income countries face in implementing measures to prevent transmission. Low income countries encounter numerous obstacles, such as inadequate infrastructure, a lack of public trust, and a high percentage of individuals working in the informal sector, who cannot work remotely from home and lose their source of income (33–35). These factors have resulted in significant challenges in controlling the virus in many low income countries, underscoring the pressing need for a global effort to address the pandemic equitably and effectively.

The focus of our research has been to explore the impact of socioeconomic determinants on the efficacy of stay-at-home measures in controlling the spread of COVID-19. By using the change in mobility as a metric for the strictness of the restrictions, without the assumption that they are an accurate quantitative representation of the real level of confinement that the population went through, we have been able to replicate the epidemic trajectory in 42 countries. Our findings indicate that a reduction in mobility is strongly associated with a decrease in virus transmission. However, we recognize that this metric may not always be an accurate representation of the true level of confinement experienced by the population. To address this, we introduced the concept of social permeability, which was estimated using Approximate Bayesian Computation. Our results suggest that low-income countries tend to have a higher permeability, indicating that restrictions were less effective in achieving an efficient population confinement.

Finally, the framework here proposed constitutes a minimal approach to capture the evolution of COVID-19 pandemic under mobility restrictions and presents different limitations. First, the model assumes a well-mixed population inside each country, neglecting possible spatial heterogeneities existing among its different regions and resulting in the aggregated values for the variables not representing a fair indicator of the evolution of the pandemic. In this case, the aggregated values for these variables do not represent a fair indicator of the evolution of the pandemic.

10.3389/fpubh.2023.1193100

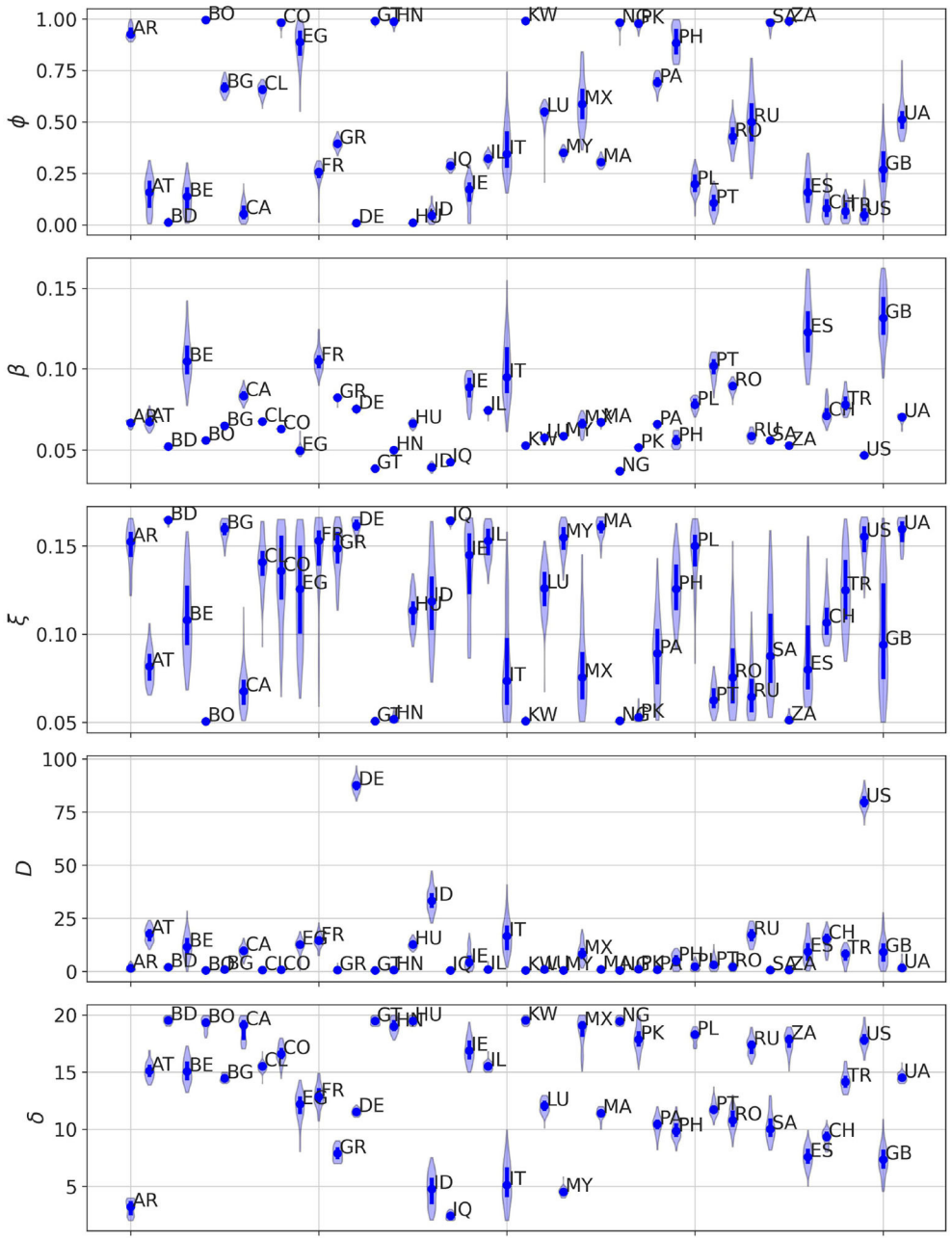


FIGURE 3
Posterior distribution for each of the free parameters of our model ($\phi, \beta, \xi, T, \delta$) obtained after calibrating the model in each of the countries here analyzed. For each parameter, dots denote the median value of the distribution whereas the solid line represents the IQR of each distribution.

In addition, our model overlooks those policies which might play an important role in bending the epidemic curves while not entailing a significant reduction in the mobility levels of the population. Some examples might be social distancing of the population, prophylaxis measures such as hands hygiene or wearing masks or ban of massive gatherings of individuals. Furthermore, mobility reduction levels were obtained from the Google COVID-19 mobility dataset, which relies on the mobility patterns estimated from smartphone users that have opted in to Google's Location History feature, which is off by default. Because of this, the results are based on the assumption that these users represent the behavior of the entire population in their respective countries. Despite all these limitations, we hope that our model paves the way to the elaboration of more sophisticated frameworks addressing the relevance of the interplay socioeconomic features and mobility reductions during epidemic outbreaks.

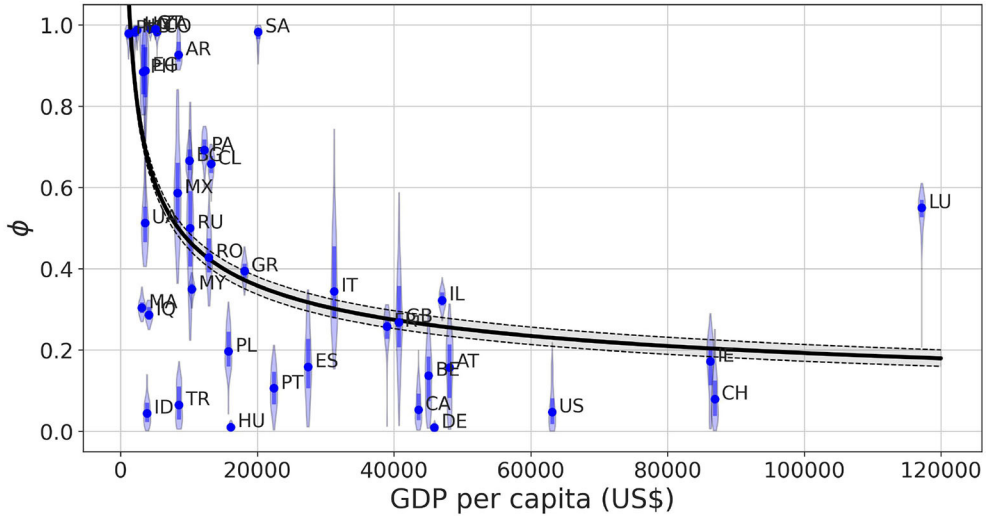


FIGURE 4
Posterior distribution obtained for the permeability parameter ϕ as a function of the GDP per capita of the country in which the model is calibrated. The shadowed region of the fit shows the 95% prediction interval of the trajectories obtained via non-linear regression $\phi(x) = ax^{-b}$, where x stands for GDP per capita and the parameters result in $a = 16 \pm 4.12$, $b = 0.39 \pm 0.03$. The solid line represents the average value of the fitted trajectories for each x. The Spearman correlation coefficient ρ_S between both variables is $\rho_S = -0.590$ with $p < 10^{-4}$.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

DS-P and JG-G conceived the research project. PV and UL performed the data analysis and calibration of the compartmental model. PV, DS-P, and JG-G drafted the manuscript. All authors contributed to the study design and methodology. All authors have read and agree with the published version of the manuscript.

Funding

PV, DS-P, and JG-G acknowledge financial support from the Departamento de Industria e Innovación del Gobierno de Aragón y Fondo Social Europeo (FENOL Group Grant E36-20R) and from Grant PID2020-113582GB-I00 funded by MCIN/AEI/10.13039/501100011033.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1193100/full#supplementary-material

References

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* (2020) 395:497–506. doi: 10.1016/S0140-6736(20)30183-5

2. Perra N. Non-pharmaceutical interventions during the COVID-19 pandemic: a review. *Phys Rep.* (2021) 913:1–52. doi: 10.1016/j.physrep.2021.02.001

3. Kapitsinis N. The underlying factors of the COVID-19 spatially uneven spread. Initial evidence from regions in nine EU countries. *Region Sci Policy Pract.* (2020) 12:1027–45. doi: 10.1111/rsp3.12340

4. Kang D, Choi H, Kim JH, Choi J. Spatial epidemic dynamics of the COVID-19 outbreak in China. *Int J Infect Dis.* (2020) 94:96–102. doi: 10.1016/j.ijid.2020.03.076

5. Raharja A, Tamara A, Kok LT. Association between ethnicity and severe COVID-19 disease: a systematic review and meta-analysis. *J Racial Ethnic Health Disparities.* (2020) 8:1563–72. doi: 10.1007/s40615-020-00921-5

6. Magesh S, John D, Li WT, Li Y, Mattingly-App A, Jain S, et al. Disparities in COVID-19 outcomes by race, ethnicity, and socioeconomic status: a systematic-review and meta-analysis. *JAMA Netw Open.* (2021) 4:e2134147. doi: 10.1001/jamanetworkopen.2021.34147

7. Benita F, Rebollar-Ruelas L, Gayton-Alfaro ED. What have we learned about socioeconomic inequalities in the spread of COVID-19? A systematic review. *Sustain Cities Soc.* (2022) 86:104158. doi: 10.1016/j.scs.2022.104158

8. Hu H, Nigmatulina K, Eckhoff P. The scaling of contact rates with population density for the infectious disease models. *Math Biosci.* (2013) 244:125–34. doi: 10.1016/j.mbs.2013.04.013

9. House T, Keeling MJ. Household structure and infectious disease transmission. *Epidemiol Infect.* (2009) 137:654–61. doi: 10.1017/S0950268808001416

10. Liu P, McQuarrie L, Song Y, Colijn C. Modelling the impact of household size distribution on the transmission dynamics of COVID-19. *J R Soc Interface.* (2021) 18:20210036. doi: 10.1098/rsif.2021.0036

11. Esseau-Thomas C, Galarraga O, Khalifa S. Epidemics, pandemics and income inequality. *Health Econ Rev.* (2022) 12:7. doi: 10.1186/s13561-022-00355-1

12. Libório MP, Ekel PY, de Abreu JF, Laudares S. Factors that most expose countries to COVID-19: a composite indicators-based approach. *GeoJournal.* (2022) 87:5435–49. doi: 10.1007/s10708-021-10557-5

13. Varkey RS, Joy J, Sarmah G, Panda PK. Socioeconomic determinants of COVID-19 in Asian countries: an empirical analysis. *J Public Aff.* (2021) 21:e2532. doi: 10.1002/pa.2532

14. Baser O. Population density index and its use for distribution of COVID-19: a case study using Turkish data. *Health Policy.* (2021) 125:148–54. doi: 10.1016/j.healthpol.2020.10.003

15. Keeling M, Rohani P. *Modeling Infectious Diseases in Humans and Animals*. Princeton, NJ: Princeton University Press (2008).

16. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput Biol.* (2017) 13:e1005697. doi: 10.1371/journal.pcbi.1005697

17. Reese H, Iuliano AD, Patel NN, Garg S, Kim L, Silk BJ, et al. Estimated incidence of coronavirus disease 2019 (COVID-19) illness and hospitalization–United States, February–September 2020. *Clin Infect Dis.* (2020) 72:e1010–7. doi: 10.1093/cid/ciaa1780

18. Iuliano AD, Chang HH, Patel NN, Threlkel R, Kniss K, Reich J, et al. Estimating under-recognized COVID-19 deaths, United States, March 2020-May 2021 using an excess mortality modelling approach. *Lancet Region Health Am.* (2021) 1:100019. doi: 10.1016/j.lana.2021.100019

19. Nations U. *World Population Prospects, The 2019 Revision - Volume I: Comprehensive Tables*. United Nations (2019). Available online at: https://www.un-ilibrary.org/content/books/9789210046428

20. United Nations. *Household Size and Composition Around the World*. Economic and Social Affairs (2017).

21. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Internal Med.* (2020) 172:577–82. doi: 10.7326/M20-0504

22. Brazeau NF, Verity R, Jenks S, Fu H, Whittaker C, Winskill P, et al. Estimating the COVID-19 infection fatality ratio accounting for seroreversion using statistical modelling. *Commun Med.* (2022) 2:54. doi: 10.1038/s43856-022-00106-7

23. Bar-On YM, Flamholz A, Phillips R, Milo R. Science Forum: SARS-CoV-2 (COVID-19) by the numbers. *eLife.* (2020) 9:e57309. doi: 10.7554/eLife.57309

24. Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C. Approximate Bayesian computation. *PLoS Comput Biol.* (2013) 9:e1002803. doi: 10.1371/journal.pcbi.1002803

25. Csilléry K, Blum MG, Gaggiotti OE, François O. Approximate Bayesian computation (ABC) in practice. *Trends Ecol Evol.* (2010) 25:410–8. doi: 10.1016/j.tree.2010.04.001

26. Gelman A. A Bayesian formulation of exploratory data analysis and goodness-of-fit testing. *Int Stat Rev.* (2003) 71:369–82. doi: 10.1111/j.1751-5823.2003.tb00203.x

27. Gutierrez E, Rubli A, Tavares T. Delays in death reports and their implications for tracking the evolution of COVID-19. *Covid Econom.* (2020) 1:116–44. Available online at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3645304

28. Arenas A, Cota W, Gómez-Gardeñes J, Gómez S, Granell C, Matamalas JT, et al. Modeling the spatiotemporal epidemic spreading of COVID-19 and the impact of mobility and social distancing interventions. *Phys Rev X.* (2020) 10:041055. doi: 10.1103/PhysRevX.10.041055

29. Van Dorn A, Cooney RE, Sabin ML. COVID-19 exacerbating inequalities in the US. *Lancet.* (2020) 395:1243–4. doi: 10.1016/S0140-6736(20)30893-X

30. Wachtler B, Michalski N, Nowossadeck E, Diercke M, Wahrendorf M, Santos-Hövener C, et al. Socioeconomic inequalities and COVID-19–A review of the current international literature. *J Health Monitor.* (2020) 5(Suppl 7):3. doi: 10.25646/7059

31. Arceo-Gomez EO, Campos-Vazquez RM, Esquivel G, Alcaraz E, Martinez LA, Lopez NG. The income gradient in COVID-19 mortality and hospitalisation: an observational study with social security administrative records in Mexico. *Lancet Region Health Am.* (2022) 6:100115. doi: 10.1016/j.lana.2021.100115

32. Drefahl S, Wallace M, Mussino E, Aradhya S, Kolk M, Brandén M, et al. A population-based cohort study of socio-demographic risk factors for COVID-19 deaths in Sweden. *Nat Commun.* (2020) 11:1–7. doi: 10.1038/s41467-020-18926-3

33. Lou J, Shen X, Niemeier D. Are stay-at-home orders more difficult to follow for low-income groups? *J Transport Geogr.* (2020) 89:102894. doi: 10.1016/j.jtrangeo.2020.102894

34. Brodeur A, Grigoryeva I, Kattan L. Stay-at-home orders, social distancing, and trust. *J Popul Econ.* (2021) 34:1321–54. doi: 10.1007/s00148-021-00848-z

35. Huang X, Lu J, Gao S, Wang S, Liu Z, Wei H. Staying at home is a privilege: evidence from fine-grained mobile phone location data in the United States during the COVID-19 pandemic. *Ann Am Assoc Geograph.* (2022) 112:286–305. doi: 10.1080/24694452.2021.1904819

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 04 August 2023
DOI 10.3389/fpubh.2023.1185928

OPEN ACCESS

EDITED BY
Pierpaolo Ferrante,
National Institute for Insurance Against Accidents at Work (INAIL), Italy

REVIEWED BY
Pedro R. Palos Sanchez,
Sevilla University, Spain
Vincenzo Auriemma,
University of Salerno, Italy

*CORRESPONDENCE
Tianyue Niu
✉ nty20@mails.tsinghua.edu.cn

RECEIVED 14 March 2023
ACCEPTED 20 July 2023
PUBLISHED 04 August 2023

CITATION
Zhang Q, Niu T, Yang J, Geng X and Lin Y (2023) A study on the emotional and attitudinal behaviors of social media users under the sudden reopening policy of the Chinese government. *Front. Public Health* 11:1185928. doi: 10.3389/fpubh.2023.1185928

A study on the emotional and attitudinal behaviors of social media users under the sudden reopening policy of the Chinese government

Qiaohe Zhang[1], Tianyue Niu[2]*, Jinhua Yang[3], Xiaochen Geng[4] and Yinhuan Lin[5]

[1]Academy of Fine Arts, Huaibei Normal University, Huaibei, China, [2]Academy of Arts & Design, Tsinghua University, Beijing, China, [3]College of Humanities, Tongji University, Shanghai, China, [4]School of Art, Nantong University, Nantong, China, [5]Xiamen Academy of Arts and Design, Fuzhou University, Xiamen, China

Introduction: Since the outbreak of the COVID-19 pandemic, the Chinese government has implemented a series of strict prevention and control policies to prevent the spread of the virus. Recently, the Chinese government suddenly changed its approach and lifted all prevention and control measures. This sudden change in policy is expected to lead to a widespread outbreak of COVID-19 in China, and the public and local governments are not adequately prepared for the unknown impact on society. The change in the "emergency" prevention and control policy provides a unique research perspective for this study.

Methods: The purpose of this study is to analyze the public's attitudes and emotional responses to COVID-19 under the sudden opening policy, identify the key factors that contribute to these attitudes and emotions, and propose solutions. In response to this sudden situation, we conducted data mining on topics and discussions related to the opening of the epidemic on Sina Weibo, collecting 125,686 interactive comments. We used artificial intelligence technology to analyze the attitudes and emotions reflected in each data point, identify the key factors that contribute to these attitudes and emotions, explore the underlying reasons, and find corresponding solutions.

Results: The results of the study show that in the face of the sudden release of the epidemic, the public mostly exhibited negative emotions and behaviors, with many people experiencing anxiety and panic. However, the gradual resumption of daily life and work has also led some people to exhibit positive attitudes.

Conclusion: The significance of this study is to help the government and institutions understand the impact of policy implementation on users, and to enable them to adjust policies in a timely manner to respond to potential social risks. The government, emergency departments, and the public can all prepare for similar situations based on the conclusions of this study.

KEYWORDS

COVID-19, deregulation of epidemic control, emotional and attitudinal behaviors, social media users, artificial intelligence, natural language processing

1. Introduction

As of September 20, 2022, the global outbreak of COVID-19 has caused 229 million cases of infection and 4.75 million deaths (1). The outbreak of the epidemic has the characteristics of suddenness, rapidity, and unpredictability. It often poses a serious threat to the life and health of the public and triggers complex emotions such as panic, dissatisfaction, and anger (2). Three years after the first COVID-19 case, China announced the end of restriction measures (later compared to the rest of the world) and experienced a large-scale spread of infections for the first time. The Chinese government announced the "Dynamic Zero COVID-19 Strategy" on January 23, 2020. On December 7, 2022, China suddenly promulgated the "The New 10 Measures" policy for epidemic prevention and control and announced the national release of the blockade (3). China's local government and the public were not prepared for the sudden change in epidemic policy. Due to the government blockade and the implementation of various security measures, the epidemic not only destroyed economic development but also changed people's lives during the pandemic. At the same time, it also has an important impact on people's mental health (4).

The public is easily influenced by various government information, and government policies are also changing. In the early days, the public obtained government information through traditional media such as newspapers and television, but lacked feedback channels. With the emergence of the Internet, the public can obtain relevant information faster, but the feedback speed is slow. Nowadays, in the era of mobile Internet, government information quickly spreads to the public (5), and the public can also make timely responses to government policies, forming a communication channel of rapid dissemination and feedback. In this case, social media has become a part of public life, and due to public concerns, the use of social media during the pandemic has become more frequent, which can not only maintain social distance but also obtain more information. The government's reputation has also shifted from eWOM to mWOM (6), and the public's feedback on government policy implementation has also changed accordingly. This is also the reason why social media users were chosen as the research object in this study, as they can provide timely data and reflect the real public attitudes.

A large amount of information that is difficult to distinguish between true and false is widely disseminated by social media, and its adverse effects are comparable to the virus epidemic, stimulating the public's nerves, stirring up social emotions, and creating extremely harmful social risks. The World Health Organization calls it "Information Epidemic". The so-called information epidemic refers to the fact that at the same time the epidemic, excessive information (some are correct and some are wrong) makes it difficult for people to find trustworthy information sources and reliable guidance (7). In the context of the information epidemic, people have a collective panic because of various information related to the epidemic, and changes in mood and attitude have seriously affected the epidemic prevention and control and social stability.

Understanding the public's emotional attitude has practical significance for current and future public health management. The public's emotions and views on the risk of the epidemic determine their personal behavior and whether they cooperate with the necessary government control measures (7, 8). Mastering the public's emotional attitude and the influencing factors will help the government in implementing effective policies to address the psychological trauma and fear experienced by the public after the COVID-19 epidemic.

As a representative Chinese online social media platform (9), Weibo is designed to facilitate interaction between users. The interaction in the Weibo environment is related to users' emotions. The communication and social value transmission brought by the social interaction of online media will allow users to experience positive emotional value changes. The close interaction of microblog users has also brought huge user value (10), user emotion and user satisfaction, and is also the embodiment of user behavior in the online social media space.

In the research of network social media users, some studies believe that posting information through social media is a behavioral plan for educating users without understanding users' needs and problems (11). However, recent studies have shown that online social media has higher user engagement than other traditional media (12). With the popularity of web social media, the Chinese government has also extended the category of government policy information used to web social platforms such as Weibo to provide information services to users (13). In emergency response, the release of government policy information mainly considers the user's adoption of the information posted by the government's official social media, and studies have been conducted on the factors that affect users' adoption of government information. Some studies propose that user communication through social media, and the adoption of government policy messages, is a value manifestation of strong democratic countries (14). In a study of user behavior analysis, sentiment analysis of Twitter posted information on U.S. presidential candidates evaluates the link between user tweets and realistic elections (15). Studies of this type of user behavior have mainly focused on a small number of social media influencers' behavioral impacts. It also shows that users' emotional behavior is influenced by other users' online emotional expressions (16). Aspects of emotion impact research targeting group social users. When encountering a disaster event, extracting network social media user behaviors can help emergency responders form stronger situational awareness of the disaster area itself (17). In China, there is a strong correlation between online social user behavior and online political contention. At the individual level, the emotional response to public events not only shapes user interpretation but also significantly impacts the social behavior of network users and the construction of public discourse. From the social level, social culture not only influences users' emotional reactions but also determines the mode of government interactions with the public and the basic framework construction of Cyberpolitical contention (18).

In this context, behavioral studies of social media users provide a more comprehensive picture of the impact of the outbreak on social development. Although numerous studies have analyzed users of social media in the face of a "post-epidemic" era, especially when the Chinese government suddenly announced a release of epidemic policy (the epidemic did not end but rather expanded further). The information of social media users is characterized

by timeliness, and the research on its behavior is a key issue in this phase of this pandemic, helping the government to respond to public opinion pressure and adjust related policies in a timely manner, meanwhile, there is also relatively little research in this area.

Emotion and attitude are parts of user behavior. In the research of user's emotional attitude, the comments with obvious emotional color and emotional tendency published by users have an important impact on enterprise performance (19, 20), public policy (21), and political decision-making (22). In recent years, user emotional analysis has emerged as a novel research approach in various fields, including entrepreneurial success (23), environmental factors in the tourism industry (24), and political decision-making and planning (25). Similarly, this has led to government departments and social organizations needing to obtain effective social feedback by analyzing and mining user emotional data, providing reliable research basis for government political decision-making and planning. Sentiment analysis on online social media extracts valuable information from natural language text to provide decision-makers with structured and actionable knowledge (26). In the process of spreading user emotions, emotional user comments are forwarded more frequently and faster than neutral online user comments (27). From the perspective of emotional media communication, there is a difference between the way users communicate their emotions through online media and face-to-face communication, and the impact of information transmitted through online media on the recipient's emotions is different from face-to-face communication. The spread of emotion in the network is mainly judged by online comments. Chinese-related online reviews are user-oriented, have a wide range of influence, and can be measured (28). The machine learning method is used to extract the emotion in online comments, and the Markov blanket model is introduced to capture the emotion in the text through conditional dependence between words and keywords and high-frequency words (29). In-group comments are usually carried out through a topic, and online comments on Weibo tend to focus on its relevant "topic". The "topic" in social media will affect user behavior (30). In view of the concentration and short text characteristics of Chinese microblogging topics in this study, Chinese microblogging is realized through topic clustering and emotional analysis, as well as prediction of hot issues in microblogging and emotional analysis of user comments (31).

Users who rely on social media for COVID-19 information may experience increased anxiety symptoms and decreased trust in the information. However, this reliance does not significantly affect preventive behavior, leading to citizens confusion regarding the adoption of preventive measures. More research is needed in this field since limited studies have explored the impact of anxiety on attitudes and behavior during the COVID19 pandemic (32). Uncertainty is a common emotional response during a health crisis, such as the COVID-19 pandemic, as it arises from perceiving an invisible threat with unpredictable outcomes. Effective communication of health information can help alleviate the fear induced by uncertainty (33).

The most direct manifestation of reflection on epidemic prevention and control policies is emotional expression (34). Through the analysis of emotions, we can understand the impact of relevant policies on society. In this situation, tracking, identifying, and analyzing the evolution of public sentiment through the mining of massive social media data not only helps to identify public emotions and implement psychological counseling, but also provides computational support for grasping the trajectory of event development and improving emergency management effectiveness. Therefore, the emotional attitude in the context of information epidemics has become a new research field that is highly concerned by various sectors.

The current research aims to explore the emotional attitude of the Chinese public and its influencing factors after the opening of China's epidemic policy. A questionnaire survey is a common data collection method in this research direction (35), but due to the limited amount of data, it is difficult to grasp the research problem in a macro and in-depth way. Using the method of big data text mining, researchers can summarize and study the impact of epidemic policy on public sentiment and attitudes and the factors behind it. Text mining is a new field in user emotion research. It collects data through the network to observe human psychology and behavior (36, 37). At present, researchers have used this method to measure the public's reaction to government policies such as mental health and social prejudice (38, 39). In the aspect of emotional attitude research, we also began to use this method to carry out relevant research (40). The use of large samples in big data mining implies strong objectivity, high timeliness, and significant impact, making it a more suitable method for this study.

Due to the gradual alternation of restrictive and permissive measures, European and American public health policy research has primarily focused on assessing the effectiveness of adopted policies and their economic consequences, rather than sentiment analysis. In contrast, China government has recently triggered relevant and emotionally charged discussions on social networks, by shifting from 3 years of strict epidemic control policies to a sudden opening up policy announced within a few days (41). Previous Chinese sentiment research has predominantly examined the emotional impact of early-stage epidemic prevention and control measures (34, 42), with less emphasis on subsequent "changes" in prevention and control policies as the epidemic situation improves. This article aims to investigate the attitudes and emotional behaviors of social media users in response to the abrupt shift in epidemic prevention policies during the period of December 1st to December 20th, 2022. The primary research questions addressed are as follows:

1. What are the prevailing attitudes among social media users?
2. How do social media users express their emotions under different attitudes?
3. What factors influence these attitudes and emotions?

2. Related work

This section discusses related work in three parts: data acquisition, training of text classification models, and sentiment analysis based on media users.

2.1. Weibo data mining

Sina Weibo is one of the most popular social media applications in China. A significant amount of work has been done to study the data of Sina Weibo. For example, Samuel et al. (43) used Weibo data to analyze social media behavior and emotional changes during emergency events. Garcia and Berton (44) used the TPACK framework to explore the design and implementation of teaching by Chinese early childhood education workers during the epidemic period using Weibo data. Boon-Itt and Skunkan (45) studied people's attitudes toward wild animals on Weibo to analyze public opinions on stray cats in China. Naseem et al. (46) used Python technology to collect Weibo data containing "Shanghai" during the epidemic period to study the public's attitude toward the image of Shanghai during the COVID-19 pandemic.

2.2. Text classification model

With the explosive growth of information, the method of manually classifying data by humans has become outdated, and using machines to automate data annotation has significant significance. Currently, there are many deep learning-based methods for text classification. TextCNN (47) uses convolutional neural networks for text classification, and its network structure is relatively simple, so the number of network parameters is small, the calculation is small, and the training speed is fast. HAN (48) uses a hierarchical structure to not only calculate attention between words, but also calculate attention between sentences. When the text/document is long, it can still obtain relatively good classification results. FastText (49) uses the method of word vector to classify text, which originated from Google's work word2vec (50, 51). FastText's model is relatively simple, so its inference speed is fast, and its accuracy is also high. Subsequently, some text classification methods based on pre-trained models gradually became mainstream, and they only need a small amount of data to achieve very good classification effects when completing specific tasks.

2.3. Sentiment analysis

There is a large amount of work utilizing data on social media for sentiment analysis during the COVID-19 pandemic. Samuel et al. (43) collected COVID-19 related tweets and used naive Bayes and logistic regression classification methods for sentiment analysis. Garcia and Berton (44) used topic identification and sentiment analysis to study a large number of tweets from Brazil and the United States, two countries with high numbers of transmission and deaths during the COVID-19 pandemic, analyzing the long-term emotional trends and their relationship with published news. Boon-Itt and Skunkan (45) conducted data mining on Twitter, collecting a large number of tweets for keyword frequency analysis, emotion analysis, and topic modeling, using natural language processing methods and latent Dirichlet allocation algorithm to identify the most common Twitter topics. Naseem et al. (46) collected a large number of tweets related to the COVID-19 pandemic on Twitter, conducting a post-evaluation of the early information flow on social media during the COVID-19 pandemic, providing information for policies applicable to social platforms. The above work conducted sentiment analysis on social media users during the COVID-19 pandemic, but lacked sentiment analysis on social media users in the context of sudden changes in epidemic policies.

3. Method

We carried out a sentiment analysis conducted on the public response to the Chinese government's announcement of removing COVID-19 mobility restrictions. The analysis was based on the following steps:

1. Data collection: we acquired sample data for analysis.
2. Data pre-processing: we conducted data cleaning to improve quality of results and manual labeling to enhance the training process.
3. Data segmentation: we transformed unstructured text into structured data through segmentation.
4. Text classification: we assigned attitudes and emotions to each comment using text classifications technics.
5. Text analysis: we extracted underlying factors through text analysis methods.

Further details on these steps can be found in the following sections, and a general representation is provided in Figure 1.

3.1. Setting

On December 7, 2022, the State Council of China issued 10 epidemic prevention policies, announcing that China's 3-year epidemic began to be fully liberalized. The announcement of the State Council marks the almost complete opening of the epidemic in China in the past 3 years, ushering in a relaxed post-epidemic era! On December 13, the trip code went offline, marking the end of the 3-year epidemic prevention policy. It triggered a heated discussion on the Internet, and the public expressed their views one after another.

3.2. Data collection

Weibo is a popular social media platform that enables instant information sharing, communication, and interaction among users through various mobile terminals such as PCs and mobile phones. With its open platform architecture, Weibo provides a simple and unprecedented way for users to publish content in realtime to the public. It has transformed the way information is transmitted on the internet and facilitated the instant sharing of information. According to Weibo's financial report for the fourth quarter and the whole year of 2020 (2022?), the platform had 521 million monthly active users and an average of 225 million daily active users in December 2020 (2022?). On Weibo, users can express their opinions on different themes, which are identified by using a "#" symbol before and after the theme name. For example, themes like #We are officially moving toward the end of the epidemic life# allow users to engage in discussions related to specific topics. In our study, we analyzed the real-time hot search list on Sina Weibo and

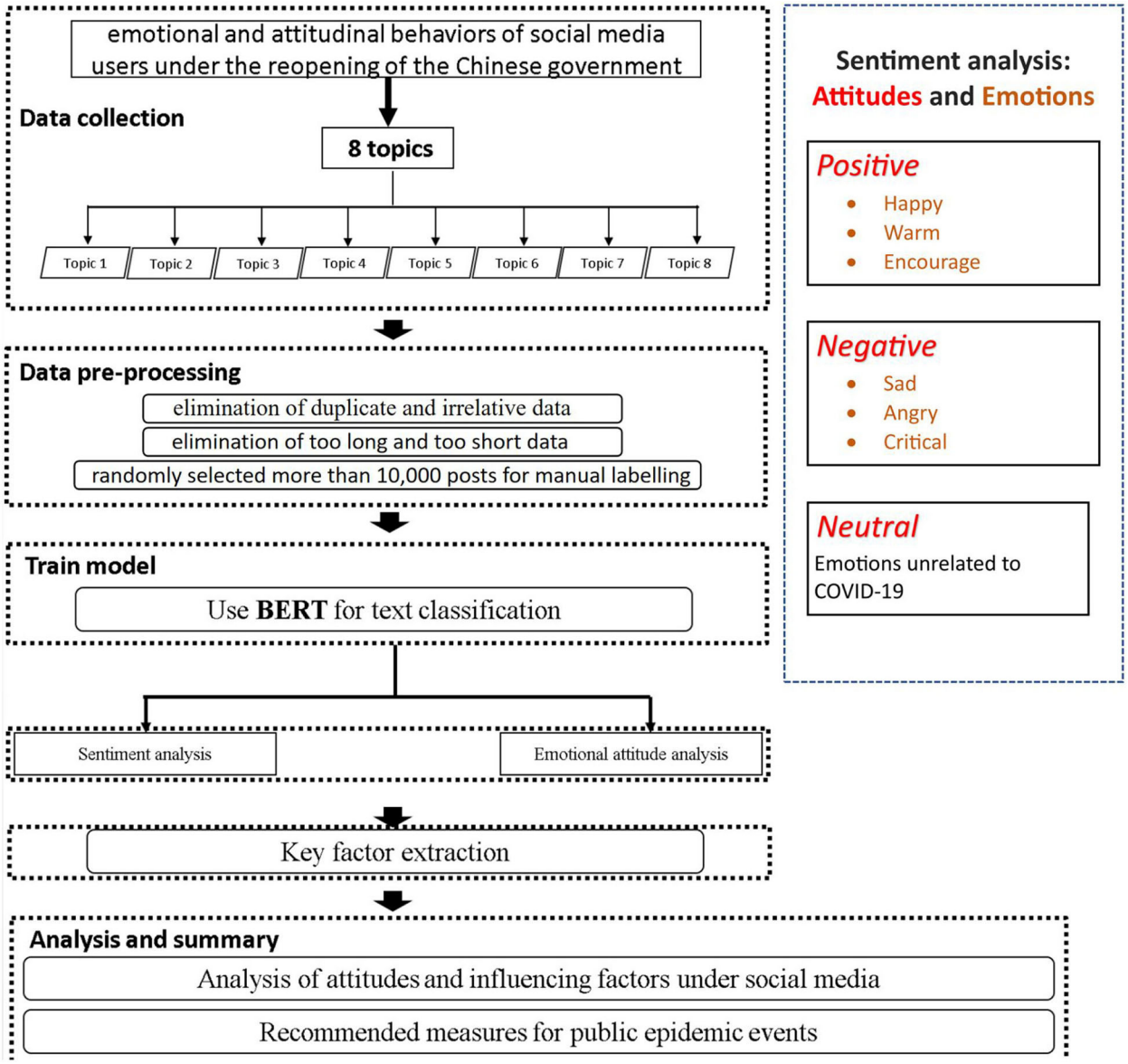


FIGURE 1
Research framework and methods of social media users' emotions and attitudes.

selected eight themes related to the epidemic with high discussion rates for data collection. Using Python technology, we collected a total of 125,686 online comments from December 1, 2022, to December 20, 2022 (as shown in Table 1). Each comment includes various attributes such as ID, BID, user ID, user nickname, Weibo text, headline article URL, publishing location, mentions (user), topic, reposts, comments, likes, publishing time, publishing tool, Weibo picture URL, Weibo video URL, and retweet ID.

3.3. Data preprocessing

To ensure data quality, we performed data cleaning and sorting on the obtained dataset. We retained only the text of the microblogs, excluding posts with fewer than 10 or more than 100 words, as well as removing duplicates. Additionally, for improved machine learning in sentiment analysis, we randomly selected over 10,000 posts related to the "release epidemic control" policy for manual labeling. Regarding the attitude reflected in each post, we utilized three labels: "negative," "positive," and "neutral." Posts unrelated to the epidemic were marked as "neutral." To capture the emotions expressed in the posts, we further subdivided the "positive" and "negative" attitudes into three types of emotions each. Specifically, the "positive" attitude encompassed emotions such as "happy," "warm," and "encouraging," while the "negative" attitude included emotions like "sad," "angry," and "critical." An example of the annotated data is shown in Figure 2.

TABLE 1 Characteristics of selected Weibo topics on China's opening up after the government announcement.

No.	Topic(#...#)	Date	Comments	Likes
1	Twenty measures to optimize prevention and control work	2022-11-10	21,593	63,984
2	Close contacts who no longer judge close contacts	2022-11-11	15,042	62,744
3	Criteria for delimitation and removal of epidemic risk areas	2022-11-21	10,080	32,253
4	Do not check the health code except for special places such as schools	2022-12-07	17,587	56,101
5	Farewell to health code	2022-12-08	13,073	24,870
6	Notice on Further Optimizing Epidemic Prevention and Control	2022-12-09	15,817	45,970
7	The new ten items of epidemic situation	2022-12-10	19,123	31,669
8	It is expected to reach the peak of infection within 1 month	2022-12-11	13,371	27,373

3.4. Segmentation

Segmentation plays a crucial role in natural language processing as it transforms unstructured textual data into structured data, establishing a standardized representation. This process involves breaking down the text into smaller units, such as words or tokens, enabling more granular analysis. By segmenting and structuring the text data, it becomes easier to perform various analyses, including text classification and text analysis. In our text classification model, we employ the trained Chinese word segmentation provided by RoBERTa as our segmentation tool. This allows us to leverage the capabilities of RoBERTa in segmenting Chinese text effectively. Additionally, for constructing the word cloud map using Sina Weibo data, we utilize the jieba thesaurus. Jieba is an excellent third-party Chinese word segmentation library in Python, utilizing a Chinese thesaurus to calculate association probabilities between Chinese characters. This enables the formation of phrases with high association probabilities. The combination of RoBERTa and jieba enables us to effectively segment the Chinese text, facilitating subsequent analyses and providing valuable insights.

3.5. Text classification

We employ text classification as a method to determine the attitudes and emotions expressed by the public in each post. Text classification involves automatically categorizing text data according to predefined classification rules or standards. It typically consists of two steps: building the feature representation of the text and training the classification model. In this study, we adopt the widely used pre-training+fine-tuning method in the field of natural language processing. Pre-training involves training the model on a large corpus of unlabeled text data, using two specific tasks: Next Sentence Prediction (NSP) and Masked Language Modeling (Mask LM). In the NSP task, the model is trained to predict whether the second sentence follows the first sentence in the original text, enabling it to capture sentence-level relationships and understand semantics. In the Mask LM task, certain words in each sentence are randomly masked, and the model predicts these hidden words based on the context provided by the remaining words. By comparing the predicted results with the original text, the model learns to fill in the masked words effectively. After completing the pre-training stage, the model can be fine-tuned to adapt to specific characteristics and requirements. This process involves using a small portion of annotated data from the target task domain, such as Multi-Genre Natural Language Inference (MNLI), Named Entity Recognition (NER), and Stanford Question Answering Dataset (SQuAD), to achieve optimal performance. MNLI assesses the model's ability to determine the relationship between two sentences (entailment, contradiction, or neutral), while NER focuses on identifying and classifying named entities (e.g., people, organizations, locations) in text. SQuAD evaluates the model's ability to generate precise answers to questions based on given passages of text. Prominent examples of this method include Google's BERT algorithm (52) and OPENAI's GPT series (53–55). Considering the advantages of BERT and its variants over other models, as discussed by (56), we utilize the BERT-based enhancement algorithm RoBERTA (57). RoBERTA utilizes longer training times, larger batch sizes, and more data to achieve improved training results. The training method of the model is illustrated in Figure 3, where the model is first pre-trained and then fine-tuned in the downstream task. We separately trained the attitude analysis model and the emotion analysis model using their respective attitude and emotion labels.

3.6. Text analysis: factor extraction

Once we have trained the attitude analysis model and the emotion analysis model, we can proceed with classifying all the data. As the classification of positive and negative attitudes encompasses three distinct emotions, we adopt a method that involves analyzing each emotion separately and then summarizing the results. For the data classified by the models, we summarize the information associated with each specific emotion. Subsequently, we perform text analysis on the collected data, generating word cloud maps that 8 represent the corresponding attitudes and emotions. From these word cloud maps, we extract the words with higher frequencies, enabling us to identify the main influencing factors for each sentiment (see Figure 4). By calculating the proportion of each factor within the original data, we can summarize the key influencing factors that contribute to a particular attitude or emotion. These factors are further classified into categories (refer to Figures 5–10), providing a comprehensive understanding of the underlying factors associated with different emotional responses.

10.3389/fpubh.2023.1185928

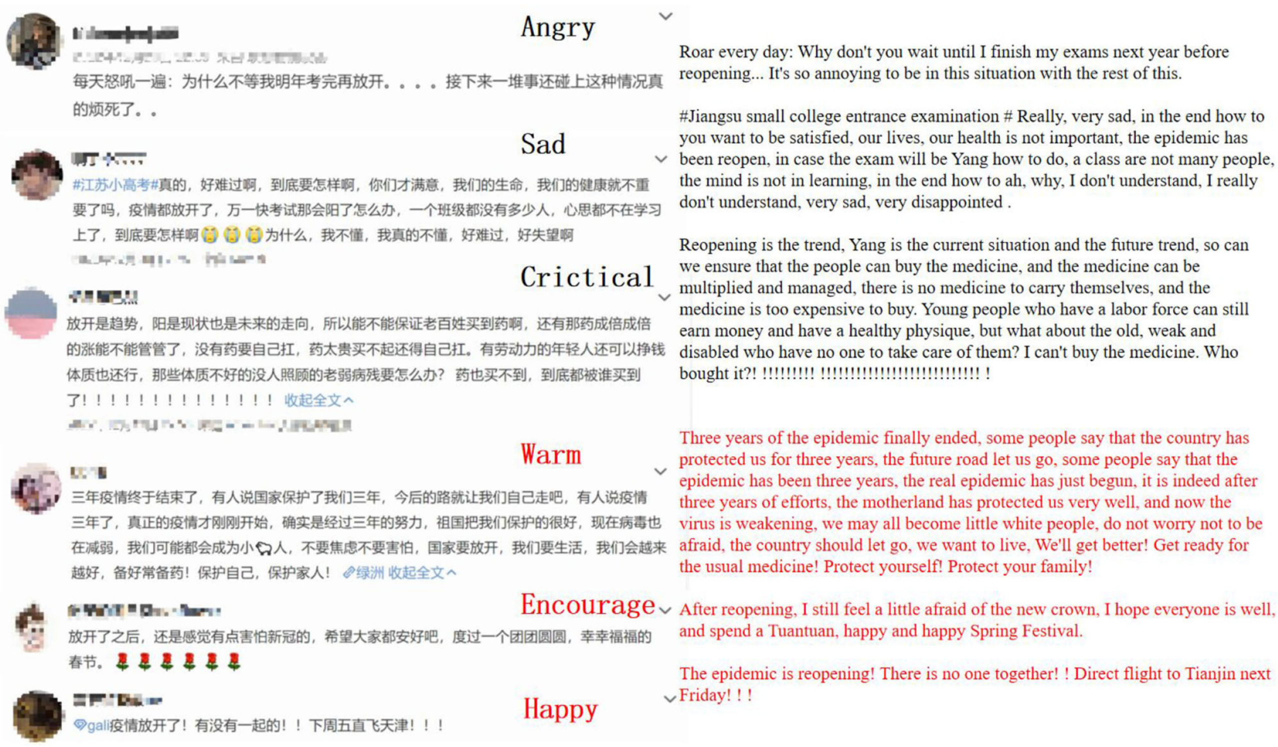


FIGURE 2
Example of performed data labeling.

4. Results and discussion

4.1. Positive attitude analysis

Positive attitude includes three emotions: "happy", "encourage", and "warm". We analyze the influencing factors of the three emotions respectively, and then summarize the influencing factors of positive attitude.

4.1.1. "Happy" emotional analysis

We summarize all the data classified as "happy" emotions, conduct text analysis to get the key influencing factors, and then count the proportion of each key factor in all the "happy" data, as shown in Figure 5. In the text analysis of "happy" emotion data, we select "concert (0.058)", "go out to play (0.032)", "film (0.017)", "go home (0.032)", "return to China (0.013)", "freedom (0.036)", "Hong Kong (0.013)", "Shenzhen (0.026)", "Guangzhou (0.067)", "travel (0.045)", "have a meal (0.009)", "at home (0.013)", "take-out (0.013)", "sit-down (0.015)", "restore (0.028)", "development (0.006)", "work (0.024)", "consumption (0.009)", "good news (0.009)", and "opportunity (0.015)", and according to the characteristics of these factors, they are classified into five categories. "Entertainment (0.107)", "family (0.045)", "getting around (0.187)", "live (0.050)", and "economy (0.091)", "Entertainment" is reflected in the public's expression of the various entertainment activities they want to carry out after the epidemic was released. Compared with the limited entertainment activities

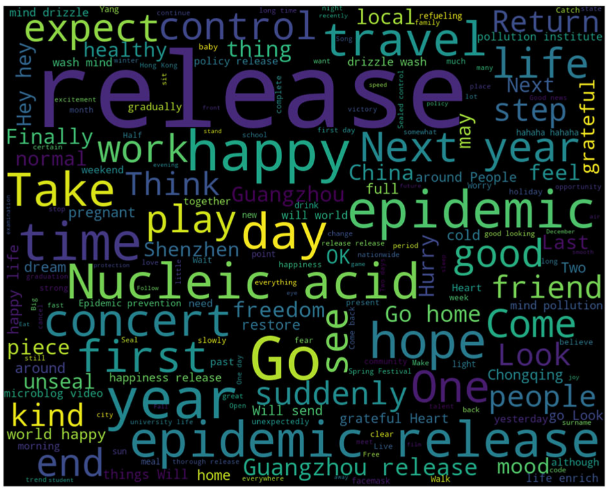


FIGURE 3
Training method of model.

during the previous epidemic control, this change makes them happy. "Family" reflect that people who leave home to work and study after the epidemic situation is released can finally feel happy to return to their families without considering being isolated. "Getting around" is reflected in the fact that people can travel freely in various cities without checking nucleic acid and travel codes after the epidemic situation is released. "Life" reflects that

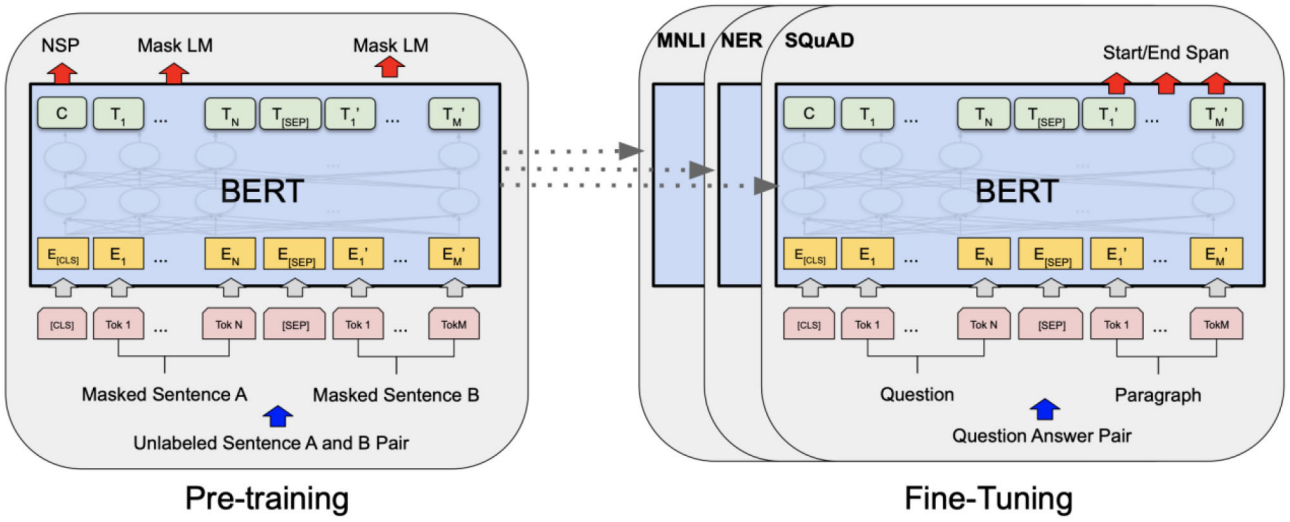


FIGURE 4
Example of word cloud.

people's life finally returns to normal, they can freely order take-out food or go out to eat, and their life gradually returns to the state before the epidemic. The economy is reflected in the fact that after the epidemic is closed and controlled, those industries limited by the epidemic finally have the opportunity to develop, and the economic situation of the country will gradually improve so that people can find their desired jobs. The proportion of "travel" is the highest. Under the policy of opening up the epidemic, the public can travel freely, which is the primary factor that makes them happy. Secondly, it is also an important factor of "happy" emotion to be able to participate in recreational activities as much as possible and the economic form will be improved.

4.1.2. "Encourage" emotional analysis

In the same way, we also conduct text analysis on all the data of encouraging emotion. As shown in Figure 6, we choose "healthy (0.073)", "protection (0.068)", "Be safe and sound (0.017)", "exercise (0.022)", "restore (0.022)", "economy (0.027)", "believe (0.017)", "academician (0.007)", "country (0.043)", "policy (0.033)", "go out (0.038)", "life (0.110)", "facemask (0.062)", "enjoy (0.010)", "freedom (0.018)", "at home (0.015)", "family (0.027)". The 18 key factors of "happy" are classified into five categories: "personal physical condition (0.180)", "national development (0.066)", "authoritative interpretation (0.083)", "life recovery (0.238)", and "home epidemic prevention (0.042)". "Authoritative interpretation" reflects that the public encourages the liberalization of the COVID-19 because the authoritative interpretation of the country and academicians shows that the liberalization of the epidemic is not a choice of lying flat, but has been overcome. From it, we can see that the proportion of "returning to life before the epidemic" is the highest. After the epidemic opens, life will slowly return to life before the epidemic. This is the primary factor of "encourage" emotion, and "personal physical condition" is also an important factor of "encourage" emotion.

#HAPPY#			
CATEGORY	FACTOR	RATE	RATE
Entertainment	concert	0.058	0.107
	Go out to play	0.032	
	film	0.017	
Family	Go home	0.032	0.045
	Return to China	0.013	
Getting around	freedom	0.036	0.187
	Hong Kong	0.013	
	Shenzhen	0.026	
	Guangzhou	0.067	
	travel	0.045	
Live	Have a meal	0.009	0.050
	At home	0.013	
	take-out	0.013	
	sit-down	0.015	
Economy	restore	0.028	0.091
	development	0.006	
	work	0.024	
	consumption	0.009	
	Good news	0.009	
	Opportunity	0.015	

FIGURE 5
Data analysis results of "happy" emotion.

4.1.3. "Warm" emotional analysis

The results of "warm" emotional data text analysis are shown in Figure 7. We have selected "shelter (0.013)", "nucleic acid

#ENCOURAGE#			
CATEGORY	FACTOR	RATE	RATE
Personal physical condition	healthy	0.073	0.180
	protection	0.068	
	Be safe and sound	0.017	
	exercise	0.022	
National development	restore	0.022	0.066
	economy	0.027	
	believe	0.017	
Authoritative interpretation	academician	0.007	0.083
	country	0.043	
	policy	0.033	
Life recovery	Go out	0.038	0.238
	life	0.110	
	facemask	0.062	
	enjoy	0.010	
	freedom	0.018	
Home epidemic prevention	At home	0.015	0.042
	Family	0.027	

FIGURE 6
Data analysis results of "encourage" emotion.

#WARM#			
CATEGORY	FACTOR	RATE	RATE
Liberalizing measures	shelter	0.013	0.213
	Nucleic acid	0.096	
	Epidemic prevention	0.042	
	Sealed control	0.036	
	epidemic control	0.012	
	clear	0.014	
Live	Go home	0.029	0.098
	Go out	0.023	
	place	0.010	
	freedom	0.025	
	Go to work	0.011	
Medical care and self-protection	facemask	0.042	0.079
	N95	0.004	
	medicine	0.009	
	hospital	0.012	
	vaccine	0.012	
Economy	restore	0.023	0.031
	consumption	0.008	
Entertainment	travel	0.018	0.098
	Guangzhou	0.037	
	Chongqing	0.014	
	concert	0.018	
	Shenzhen	0.011	
No more fear of COVID-19	Catch a cold	0.020	0.076
	immunity	0.013	
	healthy	0.031	
	Safety	0.012	

FIGURE 7
Data analysis results of "warm" emotion.

(0.096)", "epidemic prevention (0.042)", "sealed control (0.036)", "epidemic prevention and control (0.012)", "clear (0.014)", "go home (0.029)", "go out (0.023)", "place (0.010)", "freedom (0.025)", "go to work (0.011)", "facemask (0.025)", "n95 (0.004)", "medicine (0.009)", "hospital (0.012)", "vaccine (0.012)", "restore (0.023)", "consumption (0.008)", "travel (0.018)", "Guangzhou (0.037)", "Chongqing (0.014)", "concert (0.018)", "Shenzhen (0.011)", "Catch a cold (0.020)", "immunity (0.013)", "healthy (0.031)", "safety (0.012)". The 27 key factors are classified into six categories: "liberalizing measures (0.213)", "live (0.098)", "medical care and self-protection (0.079)", "economy (0.031)", "entertainment (0.098)", and "no more fear of COVID-19 (0.076)". "Liberalizing measures" is reflected in the improvement of previous measures related to epidemic control after the epidemic control. "live" is reflected in that people's encouraged be restored to before the epidemic, and everyone can travel and live freely. "Medical care and self-protection" is reflected in the fact that people can freely purchase relevant medical supplies after the epidemic situation is released, and there is no need to buy cold medicine as before. "No more fear of COVID-19" is reflected in people's understanding that the current COVID-19 is just a serious cold, no longer afraid of COVID-19, more passionate embrace of life. It can be seen that "measures related to the relaxation of the epidemic" is the primary factor that leads to the public's "enthusiasm". Under the open state of the epidemic, restrictions such as nucleic acid testing and travel codes are imposed on people, which is why the public supports the opening up of the epidemic with enthusiasm. Secondly, "life" and "entertainment" are also important influencing factors. In the open state of the epidemic situation, life will slowly recover, People can also engage in their favorite entertainment activities.

4.2. Negative attitude analysis

Negative attitudes include three emotions: "sad", "angry", and "critical". We analyze the influencing factors of the three emotions respectively and then summarize the influencing factors of negative attitudes.

4.2.1. "Sad" emotional analysis

The text analysis results of "sad" emotional data are shown in Figure 8. We selected "fever (0.026)", "serious (0.025)", "infection (0.053)", "COVID-19 (0.044)", "suffer (0.024)", "take part in the postgraduate entrance examination (0.017)", "examination (0.015)", "university (0.018)", "school (0.023)", "express (0.026)", "deliver goods (0.005)", "hospital (0.030)", "febrifuge (0.012)",

#SAD#			
CATEGORY	FACTOR	RATE	RATE
Personal health	fever	0.026	0.172
	serious	0.025	
	infection	0.053	
	COVID-19	0.044	
	suffer	0.024	
Campus life	take part in the postgraduate entrance exams	0.017	0.073
	examination	0.015	
	university	0.018	
	school	0.023	
Daily necessities	express	0.026	0.031
	Deliver goods	0.005	
Medical supplies	hospital	0.030	0.083
	febrifuge	0.012	
	facemask	0.029	
	vaccine	0.012	
Daily life	Go to work	0.019	0.076
	customer	0.018	
	colleague	0.039	
Family health	mother	0.013	0.088
	child	0.028	
	Old man	0.032	
	Family	0.015	

FIGURE 8
Data analysis results of "sad" emotion.

#ANGRY#			
CATEGORY	FACTOR	RATE	RATE
Be ill	suffer	0.025	0.177
	facemask	0.035	
	Nucleic acid	0.068	
	virus	0.029	
	drug	0.011	
	sequelae	0.009	
Live	Materials and materials	0.010	0.045
	express	0.035	
Family	At home	0.030	0.051
	child	0.021	
Government containment policy	school	0.021	0.090
	examination	0.022	
	take part in the postgraduate entrance exams	0.023	
	student	0.024	
Economy	Go to work	0.030	0.060
	economy	0.013	
	unit	0.009	
	Make money	0.008	

FIGURE 9
Data analysis results of "angry" emotion.

"facemask (0.029)", "vaccine (0.012)", "go to work (0.019)", "customer (0.018)", "colleague (0.039)", "mother (0.013)", "child (0.028)", "old man (0.032)", "family (0.015)". The 22 key factors of are classified into six categories: "personal health (0.172)", "campus life (0.073)", "daily necessities (0.031)", "medical supplies (0.083)", "daily life (0.076)", and "family health (0.088)". We can see that "personal health" is the primary factor leading to "sad" emotion, and "family health" and "medical supplies" are also important factors. In the open state of the epidemic, most people are infected, and their illness makes them feel sad, and many people can not buy medical supplies.

4.2.2. "Angry" emotional analysis

The text analysis results of "Angry" emotional data are shown in Figure 9. We selected 18 key factors, including "suffer (0.025)", "facemask (0.035)", "nucleic acid (0.068)", "virus (0.029)", "drug (0.011)", "sequelae (0.009)", "materials (0.010)", "express (0.035)", "at home (0.030)", "children (0.021)", "school (0.021)", "examination (0.022)", "take part in the postgraduate entrance examination (0.023)", "student (0.024)", "go to work (0.030)", "economy (0.013)", "unit (0.009)", and "make money (0.008)", They are classified into five categories: "be ill (0.177)", "live (0.045)", "family (0.051)", "government containment policy (0.090)", and "economy(0.060)". It can be seen that "be ill" is the primary factor leading to "angry" emotion. In the open state of the epidemic, most people are infected, resulting in their anger at the open policy. In addition, the "government containment policy" is also an important factor in the "angry" emotion. Many students are angry that the epidemic is open before the postgraduate entrance examination, which affects their examination and postgraduate entrance examination status.

4.2.3. "Critical" emotional analysis

The text analysis results of "critical" emotional data are shown in Figure 10. We have selected "infection (0.059)", "virus (0.031)", "COVID-19 (0.048)", "positive (0.019)", "fever (0.027)", "sealed control (0.048)", "control (0.026)", "isolation (0.022)", "drugs (0.012)", "N95 (0.006)", "medicine (0.011)", "febrifuge (0.011)", "vaccine (0.015)", "facemask (0.036)", "hospital (0.031)", "economy (0.019)", "lie flat (0.023)", "company (0.008)", "school (0.017)", "colleague (0.014)", "at home (0.014)", "go to work (0.025)", "life(0.020)", "abroad (0.009)", "the United States (0.010)", "domestic (0.008)", "everywhere (0.010)", and "China (0.022)". The 28 key factors are classified into six categories: "be ill (0.184)", "action (0.096)", "medical supplies (0.122)", "economy (0.050)", "live (0.090)", and "comparison at home and abroad(0.059)". The "comparison at home and abroad" is reflected in the fact that after the opening of the epidemic in China, some public expressed their dissatisfaction and expressed their feelings by comparing the

#CRITICAL#			
CATEGORY	FACTOR	RATE	RATE
Be ill	infection	0.059	0.184
	virus	0.031	
	COVID-19	0.048	
	positive	0.019	
	fever	0.027	
Action	Sealed control	0.048	0.096
	Control and control	0.026	
	isolation	0.022	
Medical supplies	drug	0.012	0.122
	N95	0.006	
	medicine	0.011	
	febrifuge	0.011	
	vaccine	0.015	
	facemask	0.036	
	hospital	0.031	
Economy	economy	0.019	0.050
	Lie flat	0.023	
	company	0.008	
Live	school	0.017	0.090
	colleague	0.014	
	At home	0.014	
	Go to work	0.025	
	life	0.020	
Comparison at home and abroad	abroad	0.009	0.059
	United States	0.010	
	domestic	0.008	
	everywhere	0.010	
	China	0.022	

FIGURE 10
Data analysis results of "critical" emotion.

domestic situation with the foreign situation. It can be seen that "be ill" is also the primary factor of "critical" emotion, and "medical supplies" is also an important factor.

4.3. Discussion

Currently, research uses Weibo data to discuss the emotional attitudes of social media users and their influencing factors, which are reflected in their online comments. By collecting a large amount of data and using machine learning methods, we explored and described the emotions, attitudes, and factors of social media users toward the open policy of the epidemic. The pandemic has had an impact on the public, and emotional expression is widespread on social media (58). Furthermore, we found that the emotional attitudes of social media users are closely related to their actual needs, rather than the release of policies. Although the government has issued an open epidemic policy, discussions related to the epidemic have not ended but have become a hot topic. We discuss the main findings shown in Figure 11 in detail.

4.3.1. Attitudes of social media users

Previous research suggested that most users had a positive attitude toward the government's implementation of the open policy for the epidemic (58). However, we found that the majority of users had a negative attitude toward the policy release and did not want the government to open epidemic prevention and control policies (59). They expected the Chinese government to maintain the blockade policy of "zero cases" domestically and epidemic prevention and control measures for foreign countries. The reason for the negative attitude of citizens is that after the epidemic policy is opened, cases will rise sharply, and people feel uneasy. Our research further found that through 3 years of epidemic prevention and control policy implementation (60), life in China has returned to normal, cases have been basically eliminated, and most Chinese citizens have not been infected with COVID-19. The public has accepted and adapted to this way of life, and sudden policy changes have caused panic among the public. Therefore, on the surface, users in China and other regions are worried about the health crisis caused by the implementation of open policies and the rise in cases. However, the reasons behind this are different. This has led to negative attitudes in China mainly being reflected in sad, angry, and critical, which also reflects the public's dissatisfaction with policy changes. At the same time, we also analyzed positive attitudes, and found that users focused on the theme of "entertainment life" after the transformation of epidemic prevention and control policies in positive emotions. Therefore, the main emotions included happy, encourage, and support.

4.3.2. Emotions of social media users

Positive attitudes, users responded positively to the government's open policy. Despite the problems caused by the epidemic policy, people actively adjusted their attitudes (61), accepted the relevant policies, and enjoyed spending time with their families. This study found that during the 3 years of implementation of China's epidemic prevention and control policy, the public's daily travel was restricted, life was inconvenient, and the economy was impacted. The sudden change in epidemic policy brought hope to the public. They are more eager to return to normal life before the epidemic (41). Therefore, positive attitudes are mainly reflected in three emotions: happy, encourage, and support, which also reflect some public support for policy changes.

Negative attitudes, users are worried about health issues and question the government's open policy (62). In previous studies, due to the impact of the epidemic, the economy declined, unemployment rates rose, people's income decreased, and a lot of negative emotions were generated. The government's open epidemic policy is to ease social conflicts and increase economic income. From this perspective, policy changes will have a positive impact on the public's lives. This study found that the majority of the public are dissatisfied with the sudden opening of the epidemic policy, and they prefer to maintain the current living status. They are worried that the open epidemic policy will lead to more serious consequences and hold a pessimistic emotion toward the epidemic opening. Negative attitudes are mainly reflected in three emotions: sad, angry, and critical.

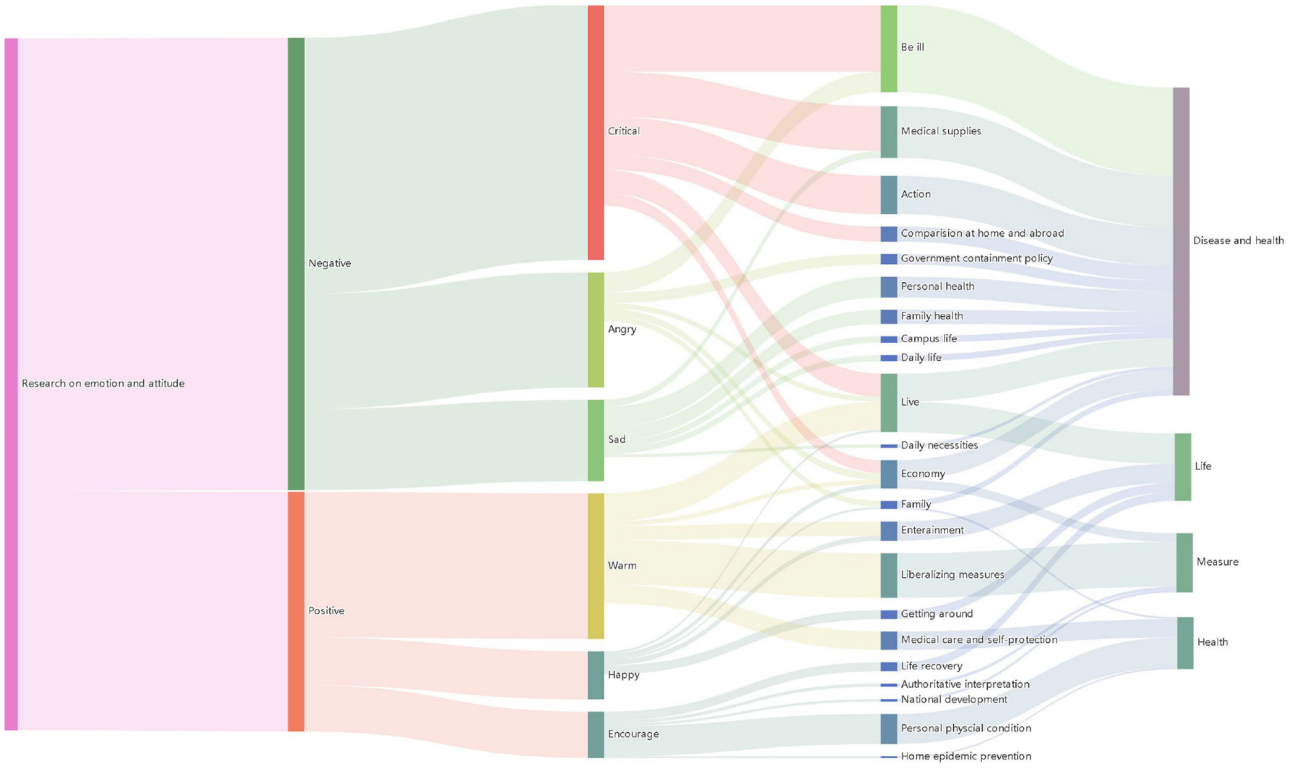


FIGURE 11
Results of the study on the emotions and attitudes of China's first opening up of epidemic control.

4.3.3. Factors influencing social media users' attitudes and emotions

In positive user attitudes, previous research has mainly discussed the overall trend of positive attitudes (63). We further studied the specific factors influencing emotions under positive attitudes, including "happy," "encourage," and "support." The main influencing factors are themes such as "freedom of movement," "return to normal life," and "government policies." Through data text analysis, the following conclusions were drawn:

In the "happy" positive emotion, the main influencing factors are the five categories of "entertainment," "family," "travel," "life," and "economy." In terms of specific expressions, "concerts" are the most discussed topic in the "entertainment" category. This also reflects that during the epidemic control period, there were fewer outdoor entertainment activities, and the public hopes for a colorful social entertainment life. "Travel" is the highest proportion of influencing factors in the user's "happiness" emotion. Due to the epidemic control policies, the government has imposed various restrictions on public travel. Faced with the open policy, travel is the public's primary choice. In the "encouragement' positive emotion, the main influencing factors are the five categories of "personal physical condition," "national development," "authoritative interpretation," "return to pre-epidemic life," and "family epidemic prevention." In terms of specific expressions, "return to pre-epidemic life" is the highest proportion of influencing factors in the public's "encouragement" positive emotion. The public's encouragement toward the open policy mainly stems from their desire for a "normal life." In the "support" positive emotion, the main influencing factors are the six categories of "life," "medical care," "economy," "entertainment," "treat it as a cold, no longer afraid," and "epidemic-related open measures." Among them, "epidemic-related open measures" are the highest proportion of influencing factors in the "support" positive emotion. The corresponding measures of the open policy are an important basis for the public's acceptance of the policy. Policy release alone cannot gain public recognition. Timely and effective response measures are helpful in gaining public "support."

In negative user attitudes, previous research has focused on analyzing the influencing factors of specific groups (64). This study found that the public's negative attitude toward the sudden opening policy during the "unexpected" epidemic mainly includes three emotions: "sad," "angry," and "critical." The influencing factors are mainly themed around "physical health" and "government policies." Through data text analysis, the following conclusions were drawn:

In the "sad" negative emotion, the six influencing factors include "personal health," "campus life," "daily necessities," "medical supplies," "daily life," and "family health." Among them, "personal health" is the highest proportion of influencing factors in the "sadness" negative emotion. After the epidemic control policy was opened, the public's biggest concern was the rapid spread of the epidemic, leading to illness. The public expressed concern about their own health status. In the "angry" negative emotion, the five main influencing factors are "illness," "life," "family," "government control policies," and "economy." "Illness" is the highest proportion of influencing factors in the "angry" negative

emotion. The Chinese public's biggest fear is getting "COVID-19," afraid of dying or suffering from sequelae due to "COVID-19," so they expressed "anger" toward the government's open policy. In the "critical" negative emotion, the six main influencing factors are "illness," "action," "medical supplies," "economy," "life," and "domestic and foreign comparisons." Similarly, "illness" is also the main influencing factor in the "critical" negative emotion. Compared with the reasons for "anger," in the "critical" negative emotion, more of the public expressed criticism of the sudden changes in government policies, and in the early stages of policy implementation, the government did not have corresponding policy plans and measures, which triggered public criticism of government actions.

Based on the above analysis, we found that the emotional attitudes of social media users are mainly based on their actual lives, which also reflects the insufficient means of the Chinese government to respond to open policies. Under sudden policy changes, social media users mainly showed negative emotional attitudes. The factors influencing negative attitudes are mainly centered around the themes of "physical health" and "government policies". After the outbreak of the epidemic, the coronavirus began to spread widely, and most people were tortured by various symptoms of the epidemic, and the supply of related medical resources was insufficient. People were plunged into an atmosphere of fear and anxiety, with the main emotions including "sad", "angry", and "critical". The analysis of the factors influencing positive attitudes is mainly centered around the themes of "freedom of travel", "returning to life", and "government policies". After the epidemic was controlled, policies related to epidemic control were gradually lifted, and people's lives could return to the way they were before the epidemic. They longed for freedom of travel and enjoyed various entertainment activities, with the main emotions including "happy", "encourage", and "support".

5. Conclusion

Previous research has revealed the necessity of conducting sociological investigations into emotions to fully understand emotions and social life. The lack of emotional communication due to the impact of quarantine policies during the pandemic has resulted in negative emotions and psychological problems. However, the factors behind emotional responses are often difficult to reflect in "small sample" data collection and are more difficult to assess from a "micro" perspective, making it difficult to make judgments at the societal level. From a neuroscientific perspective, attitudes and emotions can affect cognitive responses, and severe cases can lead to the onset of other diseases. Furthermore, understanding the physiological mechanisms behind attitudes and emotions can help in the treatment and prevention of related diseases. However, the research methods and conclusions of sociology and neuroscience seem to be unsuitable for government management in the face of effective pandemic management. Therefore, this article takes the sudden implementation of China's pandemic opening policy as the research object and uses the online comments of "Weibo" hot topics as the dataset to study social media users' attitudes, emotions, and influencing factors toward the pandemic opening policy. As is shown in Figure 11, the following conclusions were drawn through content analysis.

Firstly, in terms of social media users' attitudes, they mainly hold a negative attitude toward the "sudden" pandemic opening policy, and the proportion is quite large. The proportion of positive and neutral attitudes among the public is roughly equal. This also shows that the "sudden" pandemic opening policy is a positive measure at the national level to respond to the pandemic, but the public maintains a pessimistic view of the implementation of this policy. Secondly, in terms of social media users' emotions, positive emotions such as happiness, encouragement, and warmth express appreciation for the "sudden" opening policy. Negative emotions such as sad, angry, and critical express concerns about the "sudden" opening policy. Finally, the factors that influence the above attitudes and emotions are mainly "travel," "entertainment," "returning to normal life," "personal physical condition," and "opening measures" for positive attitudes. After classifying the influencing factors in the data, we found that the themes of "freedom of travel," "returning to normal life," and "government policies" are the main factors that affect the public's judgment of the pandemic opening policy. "Personal health," "family health," "be ill," "government quarantine policies," and "medical supply" are the main factors that affect negative emotions. After classifying the influencing factors in the data, we found that the themes of "physical health" and "government policies" are the main factors that affect the public's choice of negative emotions.

5.1. Implications

In the face of such sudden changes in epidemic prevention policies, we suggest the following:

Firstly, the public generally has a negative attitude toward "sudden" measures, and the government needs to be prepared to address this negativity. When formulating relevant policies, the government hopes to gain public acceptance and support. Therefore, in the future, when releasing similar policies, the government should prepare in advance for the supply of public life, formulate detailed measures to respond, and provide relevant health consultation services. These methods can increase the public's positive emotions and help improve their goodwill and support for the government. Secondly, "life", "health", and "measures" are the primary considerations for the public's "positive" emotional attitudes toward "sudden" policies, and the government should increase its efforts in this regard and actively respond to public demands. "Getting sick" is the most worrying negative emotional factor for the public regarding sudden government policies. At this time, the most important thing is the treatment plan when "getting sick", post-recovery body care, and issues related to life security caused by "getting sick". The government should prepare relevant handling plans in advance. Thirdly, sufficient publicity for "sudden" policies is an important means, and these factors are the main content direction of policy publicity. It is necessary to use visual poster design, new media video releases, news media announcements, and other methods to publicize relevant government measures to the public. Through these means, it is more conducive to improving the public's

emotions and attitudes toward government policies and accepting relevant policies.

In terms of emotion, the emotions ("happy", "encourage", and "warm") in a positive attitude all express appreciation for the open policy. "Travel", "entertainment", "life recovery", "personal physical condition" and "open measures" are the main influencing factors behind these emotional expressions. After classifying the influencing factors in the data, it is mainly "life", "health", and "measures" that affect the public's judgment on the open policy of the epidemic. The government hopes to be adopted and supported by the public when formulating relevant policies. Therefore, when issuing similar policies in the future, the government should ensure the supply of public life in advance, formulate detailed measures and response plans, and provide relevant health consultation services. These ways can increase the positive feelings of the public and help to improve the good feeling and warm support for the government.

5.2. Limitations and future research

There are certain limitations to this study. Firstly, the data only includes online comments on social media and has undergone text sentiment analysis, without involving other types of data such as the number of likes and shares. Combining these data with sentiment analysis of online comments can help study the spread and impact of social media users' emotional attitudes and behaviors. Secondly, although this paper analyzed users' emotional attitudes and factors toward the government's open policy, how do users' emotional attitudes change during the time of infection, illness, and recovery (usually 2–3 months)? What are the changing factors behind it? This dynamic development is also an interesting topic for research.

Data availability statement

The datasets presented in this article are not readily available because the dataset shall be applied to the research of China's open policy, and commercial use is prohibited. Requests to access the datasets should be directed to QZ zhangqh@chnu.edu.cn.

Author contributions

QZ and TN: research concept, design, and manuscript drafting. TN, QZ, JY, and XG: data collection and text classification model training. QZ, TN, JY, XG, and YL: statistical analysis, data interpretation, and review of important knowledge content. QZ and YL: obtain funding and research supervision. TN and XG: administrative, technical, and material support. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Key Project of Humanities and Social Sciences in Anhui Province's Universities, Cultural production and spiritual consumption in the post-modern perspective (sk2018A0736), and the Huaibei Normal University Top Curriculum Construction Foundation (Grant No. 2021ZLGC071).

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.

Publisher's note

References

1. Dong E, Ratcliff J, Goyea TD, Katz A, Lau R, Ng TK, et al. The Johns Hopkins University Center for Systems Science and Engineering COVID-19 Dashboard: data collection process, challenges faced, and lessons learned. *Lancet Infect Dis*. (2022) 370–6. doi: 10.1016/S1473-3099(22)00434-0

2. An L, Yu C, Lin X, Du T, Zhou L, Li G. Topical evolution patterns and temporal trends of microblogs on public health emergencies: an exploratory study of Ebola on Twitter and Weibo. *Online Inform Rev.* (2018) 42:821–46. doi: 10.1108/OIR-04-2016-0100

3. Xinhua Z. *The Chinese Government Has Announced The New 10 Measures for the Prevention and Control of the Virus, Announcing the Gradual Lifting of Containment and Opening Up the Society* (2022). Available online at: http://www.gov.cn/xinwen/2022-12/07/content_5730443.htm

4. Abbott A. COVID's mental-health toll: Scientists track surge in depression. *Nature*. (2021) 590:194–5. doi: 10.1038/d41586-021-00175-z

5. Cuillier D, Piotrowski SJ. Internet information-seeking and its relation to support for access to government records. *Government Inform Q.* (2009) 26:441–9. doi: 10.1016/j.giq.2009.03.001

6. Velicia-Martin F, Folgado-Fernandez JA, Palos-Sanchez PR, Lopez-Catalan B. mWOM business strategies: factors affecting recommendations. *J Comput Inform Syst.* (2023) 63:176–89. doi: 10.1080/08874417.2022.2041504

7. Soroya SH, Farooq A, Mahmood K, Isoaho J, Zara SE. From information seeking to information avoidance: understanding the health information behavior during a global health crisis. *Inform Process Manage.* (2021) 58:102440. doi: 10.1016/j.ipm.2020.102440

8. Abbas J, Wang D, Su Z, Ziapour A. The role of social media in the advent of COVID-19 pandemic: crisis management, mental health challenges and implications. *Risk Manage Healthcare Policy.* (2021) 14:1917–32. doi: 10.2147/RMHP.S284313

9. Zhao L, Lu Y. Enhancing perceived interactivity through network externalities: an empirical study on micro-blogging service satisfaction and continuance intention. *Decis Supp Syst.* (2012) 53:825–34. doi: 10.1016/j.dss.2012.05.019

10. Tajudeen FP, Jaafar NI, Ainin S. Understanding the impact of social media usage among organizations. *Inform Manage.* (2018) 55:308–21. doi: 10.1016/j.im.2017.08.004

11. Reddick CG, Chatfield AT, Ojo A. A social media text analytics framework for double-loop learning for citizen-centric public services: a case study of a local government Facebook use. *Government Inform Q.* (2017) 34:110–25. doi: 10.1016/j.giq.2016.11.001

12. Bonsón E, Perea D, Bednárová M. Twitter as a tool for citizen engagement: an empirical study of the Andalusian municipalities. *Government Inform Q.* (2019) 36:480–9. doi: 10.1016/j.giq.2019.03.001

13. Liang M. Current status, challenges and prospects of short video for government administration. *Electron Govern.* (2019) 199:2–10.

14. Vragov R, Kumar N. The impact of information and communication technologies on the costs of democracy. *Electron Comm Res Appl.* (2013) 12:440–8. doi: 10.1016/j.elerap.2013.06.003

15. Yaqub U, Chun SA, Atluri V, Vaidya J. Analysis of political discourse on twitter in the context of the 2016 US presidential elections. *Government Inform Q.* (2017) 34:613–26. doi: 10.1016/j.giq.2017.11.001

16. Kramer AD, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci USA.* (2014) 111:8788–90. doi: 10.1073/pnas.1320040111

17. Neppalli VK, Caragea C, Squicciarini A, Tapia A, Stehle S. Sentiment analysis during Hurricane Sandy in emergency response. *Int J Disaster Risk Reduct.* (2017) 21:213–22. doi: 10.1016/j.ijdrr.2016.12.011

18. Jinlin X. Emotion and network resistance mobilization – a case study based on the Shishou incident in Hubei Province. *J Public Manage.* (2012) 9:80–93, 126–7.

19. Lau RYK, Zhang W, Xu W. Parallel aspect-oriented sentiment analysis for sales forecasting with big data. *Product Operat Manage.* (2018) 27:1775–94. doi: 10.1111/poms.12737

20. Barbosa RRL, Sánchez-Alonso S, Sicilia-Urban MA. Evaluating hotels rating prediction based on sentiment analysis services. *Aslib J Inform Manage.* (2015) 67:392–407. doi: 10.1108/AJIM-01-2015-0004

21. Charalabidis Y, Loukis E. Participative public policy making through multiple social media platforms utilization. *Int J Electron Govern Res.* (2012) 8:78–97. doi: 10.4018/jegr.2012070105

22. Huang JY. Web mining for the mayoral election prediction in Taiwan. *Aslib J Inform Manage.* (2017) 69:688–701. doi: 10.1108/AJIM-02-2017-0035

23. Saura JR, Palos-Sanchez P, Rios Martin MA. Attitudes expressed in online comments about environmental factors in the tourism sector: an exploratory study. *Int J Environ Res Public Health.* (2018) 15:553. doi: 10.3390/ijerph15030553

24. Saura JR, Palos-Sanchez P, Grilo A. Detecting indicators for startup business success: sentiment analysis using text data mining. *Sustainability.* (2019) 11:917. doi: 10.3390/su11030917

25. Palos Sánchez PR, Folgado Fernández JA, Rojas Sanchez MA. Virtual Reality Technology: Analysis based on text and opinion mining. *Math Biosci Eng.* (2022) 19:7856–85. doi: 10.3934/mbe.2022367

26. Pozzi F, Fersini E, Messina E, Liu B. *Sentiment Analysis in Social Networks.* Morgan Kaufmann (2016).

27. Stieglitz S, Dang-Xuan L. Emotions and information diffusion in social media–sentiment of microblogs and sharing behavior. *J Manage Inform Syst.* (2013) 29:217–48. doi: 10.2753/MIS0742-1222290408

28. Shi-W, Gong X. A comparative study of online initial comment and online additional comment. *Manage Sci.* (2016) 29:45–58.

29. Zhang Y, Wallace B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:151003820* (2015).

30. Yang L, Sun T, Zhang M, Mei Q. We know what@ you# tag: does the dual role affect hashtag adoption? In: *Proceedings of the 21st International Conference on World Wide Web* (2012). p. 261–70.

31. He Y, Sun S, Niu F. A deep learning model of emotional semantic enhancement for weibo emotion analysis. *J Comput Sci.* (2017) 40:773–90.

32. Wong FHC, Liu T, Leung DKY, Zhang AY, Au WSH, Kwok WW, et al. Consuming information related to COVID-19 on social media among older adults and its association with anxiety, social trust in information, and COVID-safe behaviors: cross-sectional telephone survey. *J Med Internet Res.* (2021) 23:e26570. doi: 10.2196/26570

33. Tandoc EC Jr., Lee JCB. When viruses and misinformation spread: how young Singaporeans navigated uncertainty in the early stages of the COVID-19 outbreak. *N Media Soc.* (2022) 24:778–96. doi: 10.1177/1461444820968212

34. Velicia-Martin F, Cabrera-Sanchez JP, Gil-Cordero E, Palos-Sanchez PR. Researching COVID-19 tracing app acceptance: incorporating theory from the technological acceptance model. *PeerJ Comput Sci.* (2021) 7:e316. doi: 10.7717/peerj-cs.316

35. Martinelli N, Gil S, Belletier C, Chevalère J, Dezecache G, Huguet P, et al. Time and emotion during lockdown and the COVID-19 epidemic: determinants of our experience of time? *Front Psychol.* (2021) 11:616169. doi: 10.3389/fpsyg.2020.616169

36. Asur S, Huberman BA. Predicting the future with social media. In: *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.* IEEE (2010). p. 492–9.

37. Wilson RE, Gosling SD, Graham LT. A review of Facebook research in the social sciences. *Perspect Psychol Sci.* (2012) 7:203–20. doi: 10.1177/1745691612442904

38. Zhang S, Sun L, Zhang D, Li P, Liu Y, Anand A, et al. The COVID-19 pandemic and mental health concerns on Twitter in the United States. *Health Data Sci.* (2022) 2022:9758408. doi: 10.34133/2022/9758408

39. Lin H, Nalluri P, Li L, Sun Y, Zhang Y. Multiplex anti-Asian sentiment before and during the pandemic: introducing new datasets from Twitter mining. In: *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis* (2022). p. 16–24.

40. McGinnis EW, Lunna S, Berman I, Bagdon S, Lewis G, Arnold M, et al. Expecting the unexpected: predicting panic attacks from mood and Twitter. *medRxiv.* (2023) 1–25. doi: 10.1101/2023.01.26.23285057

41. Rahman MM, Ali GMN, Li XJ, Samuel J, Paul KC, Chong PH, et al. Socioeconomic factors analysis for COVID-19 US reopening sentiment with Twitter and census data. *Heliyon.* (2021) 7, e06200. doi: 10.1016/j.heliyon.2021.e06200

42. Pedrosa AL, Bitencourt L, Fróes ACF, Cazumbá MLB, Campos RGB, de Brito SBCS, et al. Emotional, behavioral, and psychological impact of the COVID-19 pandemic. *Front Psychol.* (2020) 11, 566212. doi: 10.3389/fpsyg.2020.566212

43. Samuel J, Ali GMN, Rahman MM, Esawi E, Samuel Y. Covid-19 public sentiment insights and machine learning for tweets classification. *Information.* (2020) 11:314. doi: 10.3390/info11060314

44. Garcia K, Berton L. Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Appl Soft Comput.* (2021) 101:107057. doi: 10.1016/j.asoc.2020.107057

45. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study. *JMIR Public Health Surveill.* (2020) 6:e21978. doi: 10.2196/21978

46. Naseem U, Razzak I, Khushi M, Eklund PW, Kim J. COVIDSenti: a large-scale benchmark Twitter data set for COVID-19 sentiment analysis. *IEEE Trans Comput Soc Syst.* (2021) 8:1003–15. doi: 10.1109/TCSS.2021.3051189

47. Kim Y. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:14085882* (2014). doi: 10.3115/v1/D14-1181

48. Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. Hierarchical attention networks for document classification. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* (2016). p. 1480–9.

49. Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of tricks for efficient text classification. *arXiv preprint arXiv:160701759* (2016). doi: 10.18653/v1/E17-2068

50. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: *Advances in Neural Information Processing Systems 26* (2013).

51. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:13013781* (2013).

52. Lee J, Toutanova K. Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:181004805* (2018).

53. Radford A, Narasimhan K, Salimans T, Sutskever I, et al. Improving language understanding by generative pre-training (2018).

54. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I, et al. Language models are unsupervised multitask learners. *OpenAI Blog.* (2019) 1:9.

55. Winata GI, Madotto A, Lin Z, Liu R, Yosinski J, Fung P. Language models are few-shot multilingual learners. *arXiv preprint arXiv:210907684* (2021). doi: 10.18653/v1/2021.mrl-1.1

56. Acheampong FA, Nunoo-Mensah H, Chen W. Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artif Intell Rev.* (2021) 54:5789–829. doi: 10.1007/s10462-021-09958-2

57. Zhuang L, Wayne L, Ya S, Jun Z. A robustly optimized BERT pre-training approach with post-training. In: *Proceedings of the 20th Chinese National Conference on Computational Linguistics* (2021). p. 1218–27.

58. Mamidi R, Miller M, Banerjee T, Romine W, Sheth A, et al. Identifying key topics bearing negative sentiment on Twitter: insights concerning the 2015-2016 Zika epidemic. *JMIR Public Health Surv.* (2019) 5:e11036. doi: 10.2196/11036

59. Samuel J, Rahman MM, Ali GMN, Samuel Y, Pelaez A, Chong PHJ, et al. Feeling positive about reopening? New normal scenarios from COVID-19 US reopen sentiment analytics. *IEEE Access.* (2020) 8:142173–90. doi: 10.1109/ACCESS.2020.3013933

60. Ahmed ME, Rabin MRI, Chowdhury FN. COVID-19: social media sentiment analysis on reopening. *arXiv preprint arXiv:200600804* (2020).

61. Vijay T, Chawla A, Dhanka B, Karmakar P. Sentiment analysis on COVID-19 twitter data. In: *2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*. IEEE (2020). p. 1–7.

62. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *Lancet.* (2020) 395:912–20. doi: 10.1016/S0140-6736(20)30460-8

63. Ridhwan KM, Hargreaves CA. Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore. *Int J Inform Manage Data Insights.* (2021) 1:100021. doi: 10.1016/j.jjimei.2021.100021

64. Xiong P, Ming Wk, Zhang C, Bai J, Luo C, Cao W, et al. Factors influencing mental health among Chinese medical and non-medical students in the early stage of the COVID-19 pandemic. *Front Public Health.* (2021) 9:603331. doi: 10.3389/fpubh.2021.603331

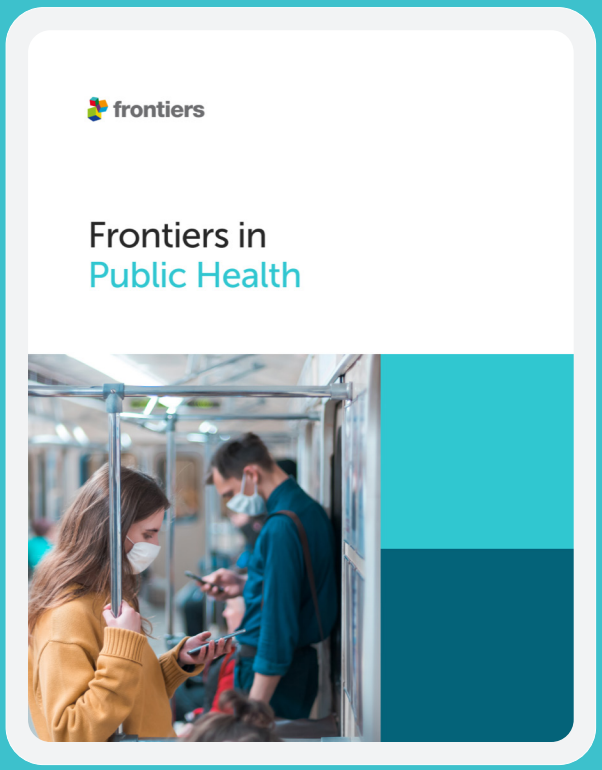
Frontiers in Public Health

Explores and addresses today's fast-moving healthcare challenges

One of the most cited journals in its field, which promotes discussion around inter-sectoral public health challenges spanning health promotion to climate change, transportation, environmental change and even species diversity.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact