



COVID-19 Vaccine Hesitancy: Analysing Twitter to Identify Barriers to Vaccination in a Low Uptake Region of the UK

Katherine Lanyi*, Rhiannon Green, Dawn Craig and Christopher Marshall

National Institute for Health Research Innovation Observatory (NIHR) Innovation Observatory, Population Health Sciences Institute, Newcastle University, Newcastle, United Kingdom

OPEN ACCESS

Edited by:

Patrick Ruch,
Geneva School of Business
Administration, Switzerland

Reviewed by:

Vasiliki Foufi,
Consultant, Geneva, Switzerland
Paris Gallos,
National and Kapodistrian University
of Athens, Greece

*Correspondence:

Katherine Lanyi
katherine.lanyi@io.nihr.ac.uk

Specialty section:

This article was submitted to
Health Informatics,
a section of the journal
Frontiers in Digital Health

Received: 29 October 2021

Accepted: 30 December 2021

Published: 24 January 2022

Citation:

Lanyi K, Green R, Craig D and
Marshall C (2022) COVID-19 Vaccine
Hesitancy: Analysing Twitter to Identify
Barriers to Vaccination in a Low
Uptake Region of the UK.
Front. Digit. Health 3:804855.
doi: 10.3389/fgth.2021.804855

To facilitate effective targeted COVID-19 vaccination strategies, it is important to understand reasons for vaccine hesitancy where uptake is low. Artificial intelligence (AI) techniques offer an opportunity for real-time analysis of public attitudes, sentiments, and key discussion topics from sources of soft-intelligence, including social media data. In this work, we explore the value of soft-intelligence, leveraged using AI, as an evidence source to support public health research. As a case study, we deployed a natural language processing (NLP) platform to rapidly identify and analyse key barriers to vaccine uptake from a collection of geo-located tweets from London, UK. We developed a search strategy to capture COVID-19 vaccine related tweets, identifying 91,473 tweets between 30 November 2020 and 15 August 2021. The platform's algorithm clustered tweets according to their topic and sentiment, from which we extracted 913 tweets from the top 12 negative sentiment topic clusters. These tweets were extracted for further qualitative analysis. We identified safety concerns; mistrust of government and pharmaceutical companies; and accessibility issues as key barriers limiting vaccine uptake. Our analysis also revealed widespread sharing of vaccine misinformation amongst Twitter users. This study further demonstrates that there is promising utility for using off-the-shelf NLP tools to leverage insights from social media data to support public health research. Future work to examine where this type of work might be integrated as part of a mixed-methods research approach to support local and national decision making is suggested.

Keywords: COVID-19 vaccination, tweets, topic clustering, artificial intelligence (AI), natural language processing (NLP), vaccine hesitancy, geo-location

INTRODUCTION

The global COVID-19 pandemic is the most significant healthcare emergency in recent memory, creating an unprecedented burden on healthcare systems (1). Since the first case was reported on 31 January 2020, there has been over 7 million cases, and to date almost 135,000 deaths associated with SARS-CoV-2 in the UK (2, 3). Globally, rapid progress was made to develop highly effective vaccines that could reduce transmission and burden of disease. At the time of data-collection, there were 3 vaccines approved for use in the UK by the Medicines and Healthcare Regulatory Agency (MHRA) (4, 5). In order to realise the potential that vaccines offer as a route out of the pandemic to a more endemic state with lower rates of severe disease, high rates of vaccination must be achieved (6). Whilst opinion varies as to the percentage uptake necessary to achieve “herd immunity,” some experts suggest it could be as high as 95% (7).

The UK vaccination programme commenced on 08 December 2020. As of 15 August 2021, every adult UK resident had been offered the first dose of a MHRA-approved vaccine. Initially, public enthusiasm to receive a COVID-19 vaccine was high, with supply—not demand—being the limiting factor (6, 8). This enthusiasm continued until the start of July 2021 when a plateau in the culminative number of vaccines administered was observed (9). As of 15 August 2021, 89.3% of those eligible in the UK had received their first dose (8). However, significant regional variation in vaccination rates was (and remains) apparent across different regions of the country. In London, for example, only 82.3% of residents eligible for a vaccine had received a first dose by the same date (August 2021) (9, 10). As demonstrated by previous immunisation campaigns for other transmissible diseases (e.g., winter flu and meningococcal serogroup C), greater uptake can be achieved through targeted intervention strategies addressing specific barriers to vaccination (11, 12). For interventions to be appropriately targeted, the reasons underpinning vaccine hesitancy in low-uptake areas must be clearly understood. Traditional methods of collecting data relating to public opinion, such as surveys or focus groups, are time, money, and resource intensive and faced particular challenges in the current pandemic climate (13). Further in a fast-moving situation, such as the pandemic, it remains unclear whether traditional methods offer a more robust solution than a more rapid, pragmatic approach to gaining insight and facilitating rapid development and delivery of interventions that are crucial (14, 15).

More than 77% of the UK population are active on social media (e.g., Twitter, Facebook, Reddit), and with usage increasing during the pandemic, vast amounts of rich data exist within these so called “soft-intelligence” sources (16, 17). Novel approaches capable of extracting and analysing these data to provide actionable insight into public perceptions have the potential to transform the landscape of public health research (15, 18, 19). In addition to the increased volume of data that soft-intelligence sources provide, user generated content on social media is freely volunteered and not restricted to the scope of the question being posed (13). Previous studies have shown that when leveraged using artificial intelligence (AI)-based techniques including NLP, insights from social media can be useful to rapidly understand public opinion, sentiment, and behaviour (20–22).

The aim of this work was to further investigate and explore the value of using insight gained from social media, as a meaningful source of intelligence to support public health research. In this article, we report the findings from a short case study, in which we deployed a NLP platform to rapidly detect and analyse key barriers to vaccine-uptake from a sample of geo-located tweets posted from users in the region of London, UK.

Abbreviations: AI, Artificial Intelligence; BCW, Behaviour Change wheel; COM-B, Capability and Behaviour Model; MHRA, Medicines and Healthcare Regulatory Agency; NIHR, National Institute for Health Research; NLP, Natural Language Processing; SAGE, Strategic Advisory Group of Experts; WHO, World Health Organisation.

METHODS

Search and Data Collection Strategy

We selected Twitter as our chosen data source for collection and analysis. Twitter is a social media platform where users can post short messages “tweets” up to 280 characters long. Twitter is amongst the most popular social media platforms in the UK with over 17.5 million active users, as of July 2021 (23). Twitter data has been used successfully in previous surveillance studies as a source of real-time user-generated data to track trends in public dialog and perceptions over time and recognise what is happening on the ground during a viral pandemic (20, 24, 25).

We developed a search strategy comprising the following list of terms:

(vaccine OR vaccines OR vaccinate OR vaccinates OR vaccination OR vaccinations OR vaccinated OR vax OR vaxx OR anti-vax OR anti-vaxx OR antivax OR antivaxx) OR (covid OR coronavirus) (moderna OR pfizer OR BioNTech OR AstraZeneca)

This strategy was used to search for any relevant geo-located tweets posted by users in London, UK relating to COVID-19 vaccines or the vaccine roll-out. The specific search terms and syntax were generated through discussion between members of the National Institute for Health Research (NIHR) Innovation Observatory soft-intelligence and information research working groups, along with scanning relevant literature, including recently published studies and news articles (25–27).

Once the search strategy was confirmed, we began prospectively searching for and collecting relevant tweets via Twitter’s advanced search application interface (28). The search ran over an 8-month period, between 30 November 2020 and 15 August 2021. This time-period covered the approval of the first vaccine in the UK through to all UK citizens aged 18 years and over being offered the first dose (9). All tweets identified in the search were anonymised to protect the privacy of users.

Data Analysis

Natural Language Processing

An advanced AI-based, text analytics platform using NLP was used to initially analyse the tweets. The analytics platform, “Wordnerds,” is described by its developers as a “text analysis and insights platform using machine learning techniques” (29). In particular, this off-the-shelf platform supports analysis of meta-data, topic, and sentiment to understand the context of a tweet and group tweets together into topic clusters that contains tweets relating to each other, or discussing similar issues. This facilitates a more accurate and sophisticated insight in to the vaccine conversation on Twitter compared to methodologies which rely solely on a qualitative count of single words, phrases, or hashtags (30).

For this study, we used the platform to analyse the volume and sentiment of the collated tweets, and to identify key topics of negative discussion within the dataset. On loading the tweets, the platform was able to determine the sentiment of each tweet and then clustered them accordingly with others that discussed the same (positive or negative) topic.

Based on the analysis of the initial corpus of tweets, the platform automatically generated 12 clustered topics of conversation related to COVID-19 vaccination with negative sentiment. The clustered topics included the following:

1) “covid vaccine,” 2) “vaccine passports,” 3) “people vaccinated,” 4) “worry vaccine,” 5) “vaccines work,” 6) “vaccine rolled,” 7) “second dose,” 8) “having vaccine,” 9) “thing vaccines,” 10) “booking vaccine,” 11) “coronavirus vaccine” and 12) “az developed vaccine.”

The tweets contained within these 12 topic clusters were used as our sample to ascertain potential barriers to vaccine hesitancy via qualitative document analysis.

Coding Tweets to Hesitancy Themes

Using the sample set generated by the platform, tweets were manually coded by 2 researchers (KL and RG), independently, to one of 6 pre-determined themes relating to vaccine hesitancy:

1) Mistrust, 2) Safety, 3) Ineffective, 4) Access, 5) Underrepresentation, or 6) Complacency.

These themes were developed based on the Scientific Advisory Group for Emergencies (SAGE) working group’s 3C model of

vaccine hesitancy, published by the World Health Organisation (31). This model establishes three core barriers that determine levels of vaccine uptake: confidence (i.e., the level of trust in the efficacy and safety of the vaccine), complacency (i.e., the perceived need for the vaccine), and convenience (i.e., how accessible the vaccine is to people) (31, 32). The 3C model emphasises that whilst all vaccine hesitancy is grounded in “the 3Cs,” the specific issues underpinning these core barriers (such as safety, efficacy, cost and trust) are context specific to a particular vaccine, and the circumstances for which it is being developed (31, 33).

Following close examination and consideration of the 3C model, and based on our study’s context and target population, we arrived at the 6 tailored themes listed above. We note that the themes we have selected do not cover all those considered by the 3C model. For example, affordability, an important reason underpinning the 3C’s model convenience barrier, was not included for this study, since vaccines are freely-available to the UK public via the National Health Service (NHS). London has a relatively high population of ethnic-minority residents who, according to polling, are least likely to get vaccinated (27, 34, 35). Therefore, we also

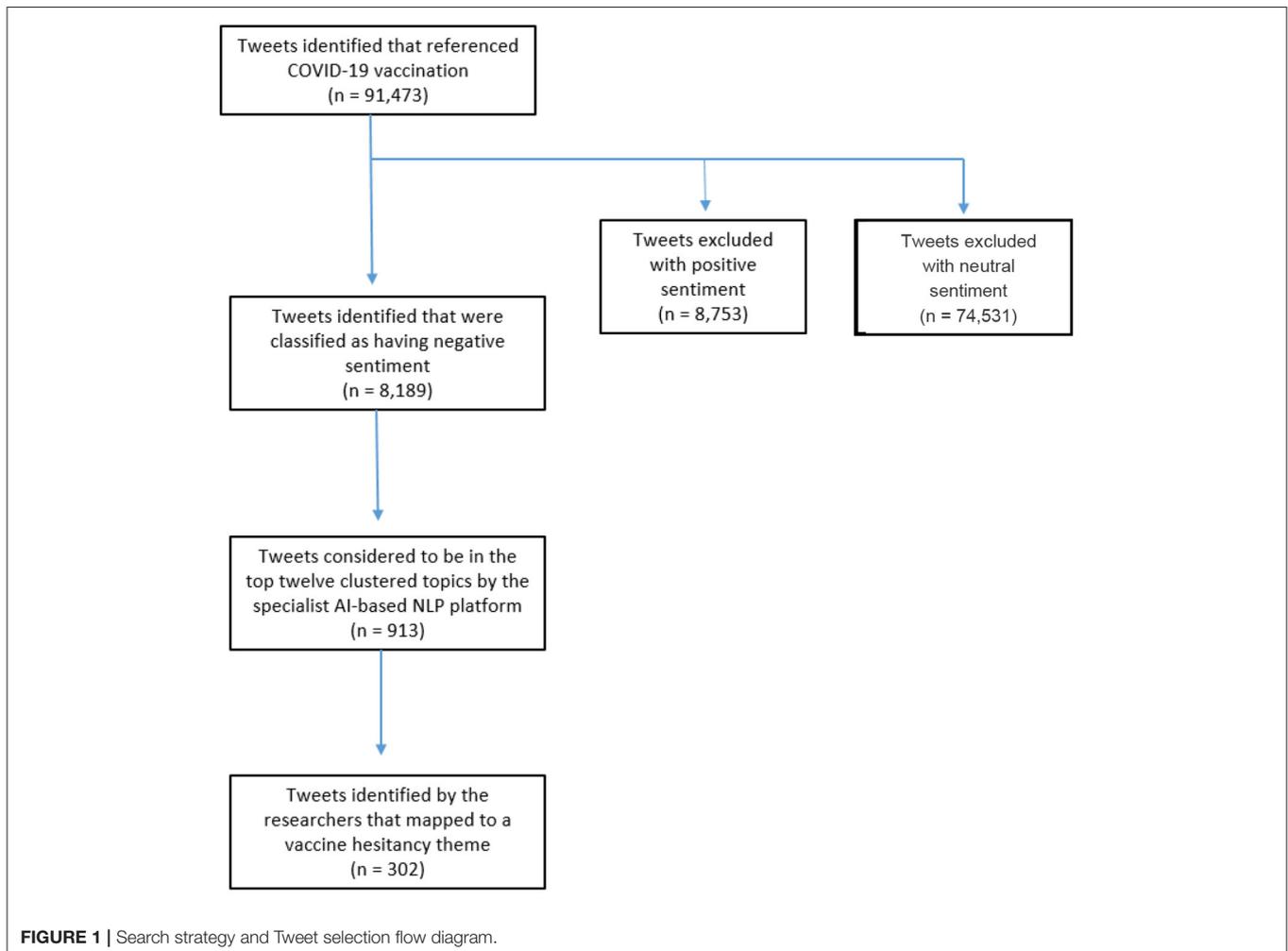


TABLE 1 | Exemplar tweets mapped to each theme.

Mapped barrier	Selected example tweets
Safety	<i>"I'm getting my Covid vaccine tomorrow however, I'm scared it'll affect fertility 😬"</i> <i>"Nothing at all to worry about... Pfizer Covid vaccine: Side effects may include pulmonary embolism "</i>
Mistrust	<i>"I'm sorry, but if @RishiSunak is found to have been making personal profits from investing in COVID vaccine companies, he needs to be sacked. Does conflict of interest have no meaning to the @Conservatives ? So, so corrupt."</i> <i>"Spot on variants and vaccine passports are the threat to pushing this vaccine, it all started off for the old and vulnerable and protecting the nhs, now cancer and heart trouble have taken over, this covid charade is a smoke screen."</i>
Access	<i>"All I want is a chance to at least book an appointment for a vaccine, but here I am trying to learn code and sign up to bots to get a slot"</i> <i>"I have received 2 messages regarding COVID vaccine but with all the scammers I can't even be sure if this is valid 😬"</i>
Ineffective	<i>"If antibody levels are this high, yet we are still seeing so many 'cases' surely this is now just endemic? The vaccines don't prevent you getting or transmitting so what's the point of vaccine passports when vaccination rates are touching 90%"</i> <i>"Why are there still more deaths in the 80+ age group than any other, given this most vulnerable group was vaccinated first? There is a clear inconsistency here, either the vaccine does not work, or COVID is not the cause of their death."</i>
Complacency	<i>"I tired hearing bout COVID, curfew, vaccines and all that other cxnt"</i> <i>"I'm not a fan of national sheep training (aka football) but vaccine passports are just another bungle during the great covid over reaction"</i>
Under-representation	<i>"Black folk still need to take the COVID vaccine but I honestly have no time for White ppl blowing the anti-vaxer horn when Black ppl share their fears. This isn't the same conversation as your middle class White mum who just discovered essential oils & herbs."</i> <i>"So I've not long found out that one of the covid vaccines has only been tested on white people. Which makes no sense as 'bame' are the more at risk group and have been neglected once again in research."</i>

government. A small number of tweets referenced conspiracy theories such as mass vaccination being a government ploy to weaken the immune system of the “sheep,” or COVID-19 pandemic being deliberately manufactured or deliberately exaggerated by pharmaceutical companies “just for profit” from vaccine administration. Further, users were sceptical of the government’s competency to deliver the rollout, citing previous failings during the pandemic, including “test and trace” and “care homes” as reasons not to trust the government. Many did not believe the existence of new “mutant strains” or “new variants,” perceiving them to be “government lies” constructed to cover-up the vaccines “never really worked at all.” This sense of mistrust was further fostered by circulation of a news story reporting that “big pharma” companies were “protected” from “being sued” or accepting any “legal liability” as a result of any adverse events that emerged as a result of the vaccine.

Tweets Mapped to the “Ineffective” Theme

There were lots of discussions surrounding vaccines being “less effective than promised” or “ineffective against variants” or how a “second dose delay” would reduce long-term immunity. However, it was difficult to infer if this discussion meant that they would be likely to refuse immunisation as a result of these concerns about effectiveness. The negative sentiment arising from tweets coded to this theme was generated by a sense of despondency that arrival of the vaccines would not resolve the crisis as hoped. Whilst the majority of posts highlighting that you could “still get COVID” despite being vaccinated were using this fact to argue against implantation of the controversial vaccination passports rather than as a reason to refuse vaccination. However, a small number of users were asking “what’s the point?” of getting an “ineffective” vaccine that did not prevent COVID transmission, implying that frequent discussion regarding the lack of vaccine effectiveness could reduce uptake.

Tweets Mapped to the “Access” Theme

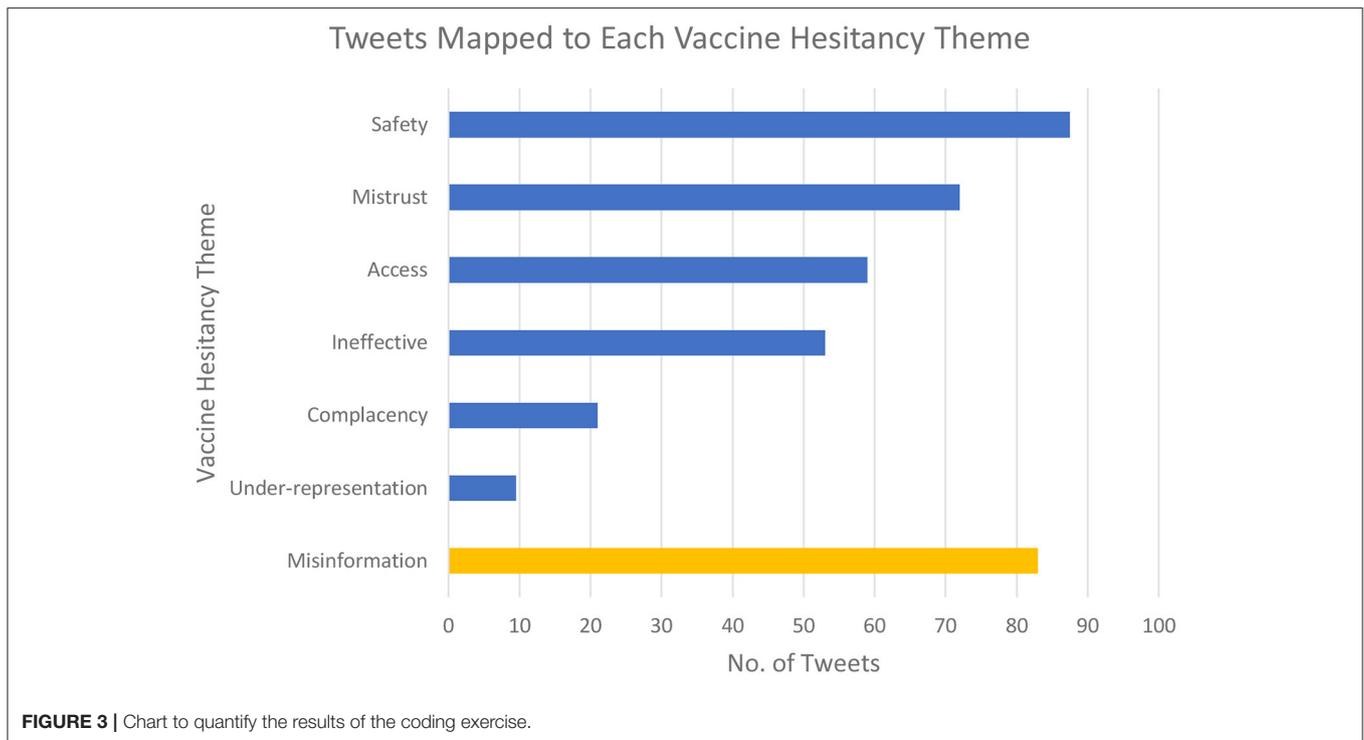
Difficulties relating to accessing vaccines were heavily discussed within our sample of tweets. However, many tweets did not reflect genuine inability to access vaccines, but rather pro-vaccine users upset by people “queue jumping,” particularly those who were not considered “vulnerable.” A key issue consistently emerging within tweets coded to access was difficulty booking an appointment with frustration expressed at how “time-consuming” or “impossible” the process was. In particular, users reported making numerous attempts to secure an appointment using the NHS online system, with some remarking that it was “harder than getting Glastonbury tickets.” A small number of tweets reported how vaccine-related “scams” and “fake texts” had made them doubt NHS text message reminder notifications, resulting in missed appointments.

Tweets Mapped to the “Complacency” Theme

Some users’ tweets expressed a complacent attitude towards the need for vaccines. Tweets using phrases like “what’s the point?” or “no need” were often posted in response to vaccine-related news articles being shared. Some people believed that only the vulnerable needed to be vaccinated and that mass-vaccination had been an “over-reaction.” Complacency detected amongst some users emerged as declarations that they were “over COVID,” and just wanted a return to normality.

Tweets Considered Misinformation

Tweets classified as misinformation were heavily biased towards safety, mistrust, and efficacy concerns. A large volume of tweets were identified as misinformation either in the form of a lack of user knowledge or posts from so called “anti-vaxxers” deliberately attempting to spread rumours to discourage vaccination. A number of individuals did not believe the coronavirus pandemic was real and many posts referenced anecdotal “evidence” of how dangerous the vaccines were such as: “HORROR! as 27 die suddenly after taking Pfizer jab” or “man left in agony as skin peels off hours after getting Astrazeneca vaccine.” Although most tweets tagged as misinformation appeared to be coming from “anti-vaxxers,” or vaccine hesitant individuals, a small minority



of users were posting inaccurate information to encourage vaccination. For example, tweets were identified that shared false statistics on the relative risk of blood clots after taking the AstraZeneca vaccine compared to taking the contraceptive pill. Social media is a vital tool for disseminating health information; however, the high proportion of tweets in our sample coded as misinformation also highlights the potential concerns. Fostering online communities who refute scientific advice and instead make healthcare decisions based on the false online information creates substantial public health risk.

DISCUSSION

The purpose of this study was to further investigate the application of NLP techniques as a means of gathering evidence from unstructured, soft-intelligence data sources and assess the utility of this to inform public health research or decision making. As a case study we deployed a commercial, AI-driven NLP platform to leverage insights from Twitter data, with the aim of rapidly identifying key barriers to COVID-19 vaccination uptake amongst users in London.

Throughout the analysis period (30 November 2020 to 15 August 2021), 91,473 Tweets referencing COVID-19 vaccines were posted from London Twitter accounts. The specialist text analytics platform we deployed assigned all of the collected tweets as having positive, neutral, or negative sentiment. The platform utilised machine learning to automatically extract the tweets from the 12 most common topic clusters underpinned with negative sentiment to generate a sample corpus of 913 tweets to perform qualitative analysis.

Results from our qualitative analysis highlighted the polarising views amongst different users in the online vaccine

discourse. We identified concerns over vaccine safety, and mistrust towards the government or pharmaceutical companies to be the two major themes relating to vaccine hesitancy. We also identified numerous tweets that contained and reported misinformation. This further highlights that whilst social media can be a powerful means to disseminate useful health-related information, it can equally be used to spread false and potentially harmful information, thus posing a public health risk.

Across all vaccine hesitancy themes the main issues preventing uptake were:

- Concern that vaccines developed so quickly must be experimental and fears that inadequate testing could result in adverse side effects amongst those who take it.
- Beliefs that pandemic was being falsely reported by the media and accusations levelled at the government of fabricating data to coerce mass vaccination.
- Anger and anxiety that pharmaceutical profits were being prioritised before population safety.
- Scepticism regarding vaccine efficacy.
- Belief that only those who were old or vulnerable needed to be vaccinated.
- Access issues. In particular, difficulty using the online booking system.

Our approach identified similar themes underpinning vaccine hesitancy compared to previous longitudinal surveys assessing vaccine hesitancy in the UK (32). However, by utilising the automated topic and sentiment clustering capabilities of the platform we deployed, a case can be made that the findings were acquired using less time and resources on the part of the researchers. Additionally, this case study used an off-the-shelf platform rather than an internally developed, bespoke, AI

tool, making this technology more accessible to the researchers. The mixed-method approach adopted allowed for a more nuanced analysis to be undertaken, using a more robust established methodology. Despite the need for human input, the overall resource required to produce this research was reduced considerably through using the NLP tool.

This study contributes to a growing body of work investigating Twitter as a source of soft-intelligence, which can be used to capture real-time public insights, attitudes, and emerging trends concerning a particular health issue (15, 18, 36). Most of this existing research has largely been conducted through qualitative analysis of a small sample of tweets that are selected randomly from within a large dataset (40, 41). Here we present a case study that helps demonstrate the advantages of using a specialist AI-driven NLP tool that can be tailored to generate a corpus of tweets capturing the most common negative topics of discussion rapidly and automatically within a large dataset as a basis for further contextual evaluation of the topics.

Using machine learning to generate a high-quality sample set of the most relevant tweets enabled a quicker and more focused qualitative analysis. It removed the need for a researcher to manually sift through many 100s and 1000s of irrelevant posts, thereby increasing the efficiency and reducing the time and resources necessary to answer the research question.

This is a small case study that demonstrates the feasibility of AI tools to efficiently compile a corpus of relevant data to be analysed more robustly using established methods. Like other research of these methods, it shows that this is a promising methodology that has the potential to become a valuable addition to a well-established portfolio of evidence synthesis methods. As with any new methodology, further research is still required, with an initial focus on the issues of generalisability and bias that the use of these types of data and tools may bring. However, our case study and other work in this field suggests that there is a place for these analysis tools, alongside more established methods of evidence synthesis, when addressing some public health research questions.

LIMITATIONS

The data that is available from social media is a sub-population and hence raises questions regarding generalisability. For our case study, the demographic of London Twitter users is not representative of the demographic of the London population and may differ in terms of age, gender, and socioeconomic status. Since it is not possible to collect such demographic data while maintaining the users' anonymity, this is a limitation faced when using this social media site as a source of soft intelligence. More work is required to fully explore this; currently it is unclear whether the issue is as significant as perceived.

As with many AI NLP solutions, the nuance of language can lead to odd results. In this example, the platform struggled to ascertain the true context of certain tweets. For example, based on the platform's automatic topic and sentiment clustering, it would have appeared that vaccine passports were amongst

the most significant barriers to vaccination. In actuality, most tweets discussing vaccine passports supported the vaccine but strongly opposed vaccine passport introduction. We tried to overcome this limitation by taking a mixed methods approach and incorporating some element of manual screening to ensure the validity of results. However, as with all research involving qualitative data analysis, there is a need to ensure that there is a consensus between reviewers and consistency in approach. Given that our search strategy generated almost 100,000 tweets, we decided to limit topic and qualitative analysis to negative sentiment tweets. Whilst tweets expressing vaccine hesitancy were most likely to have been classified as negative, there may have been some classified as neutral and therefore not included for analysis. We would suggest based on this, that it may be of benefit in future work to review samples of the tweets excluded at the neutral sentiment and topic clustering stages to further assess the viability of the platform (see **Figure 1**). There may be limitations to the methods we used for the case study itself, for example we may have expanded our search terms to include more informal words for vaccines such as "jab." However, these are all subjective decisions that like any evidence synthesis research should be determined *a priori*, transparently report and justified. Given the focus of the paper these limitations are not discussed fully.

FUTURE WORK

This case study encapsulated a 6-month period of the pandemic. To truly assess the use of the platform to provide a rapid analysis of public sentiment in the public health space, it may be of worth capturing a smaller time period with a broader geographical range. Additionally, this tool could be deployed on other public forums such as MumsNet and Redditt to assess vaccine hesitancy amongst specific population groups who have lower vaccination rates. Finally, work to explore how this kind of analysis might inform and evolve existing mixed methods approaches, including those leveraging established behavioural models (e.g., BCW and COM-B) (42, 43). This could help to develop targeted intervention strategies, that have maximum impact on vaccine uptake.

We believe that this work demonstrates the utility for off-the-shelf NLP tools to leverage insights from social media data to support public health research as part of a mixed-methods approach or during times of crisis when rapid and reactive public health engagement is needed.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

KL and CM contributed to conception and design of the study. KL and RG performed the mapping and qualitative document analysis. KL wrote the first draft of the manuscript. KL, CM,

RG, and DC wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This project was funded by the National Institute for Health Research (NIHR) (HSR1C-2015-1009/Innovation Observatory).

REFERENCES

- Bonell C, Michie S, Reicher S, West R, Bear L, Yardley L, et al. Harnessing behavioural science in public health campaigns to maintain 'social distancing' in response to the COVID-19 pandemic: key principles. *J Epidemiol Community Health*. (2020) 74:617–9.
- GOV.UK. *COVID-19 Cases in the United Kingdom*. (2021). Available online at: <https://coronavirus.data.gov.uk/details/cases> (accessed September 09, 2021).
- GOV.UK. *COVID-19 Deaths in United Kingdom*. (2021). Available online at: <https://coronavirus.data.gov.uk/details/deaths> (accessed September 15, 2021).
- Public Health England. *COVID-19 Vaccination Programme: Information for Healthcare Practitioners*. (2021). Available online at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1009174/COVID-19_vaccination_programme_guidance_for_healthcare_workers_6_August_2021_v3.10.pdf (accessed September 15, 2021).
- World Health Organisation (WHO). *COVID-19 Vaccine Tracker and Landscape*. (2021). Available online at: <https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines> (accessed September 15, 2021).
- Department of Health & Social Care. *UK COVID-19 Vaccines Delivery Plan*. (2021). Available online at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/951928/uk-covid-19-vaccines-delivery-plan-final.pdf (accessed September 16, 2021).
- World Health Organisation (WHO). *Coronavirus Disease (COVID-19): Herd Immunity, Lockdowns and COVID-19*. (2020). Available online at: <https://www.who.int/news-room/q-a-detail/herd-immunity-lockdowns-and-covid-19> (accessed September 15, 2021).
- GOV.UK. *All Young People Aged 16 and 17 in England to Be Offered Vaccine by Next Week*. (2021). Available online at: <https://www.gov.uk/government/news/all-young-people-aged-16-and-17-in-england-to-be-offered-vaccine-by-next-week> (accessed October 17, 2021).
- GOV.UK. *Vaccinations in United Kingdom*. (2021). Available online at: <https://coronavirus.data.gov.uk/details/vaccinations> (accessed September 15, 2021).
- Office for National Statistics (ONS). *Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland: 2021 Local Authority Boundaries*. (2021). Available online at: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalescotlandandnorthernireland> (accessed November 26, 2021).
- Dexter LJ, Teare MD, Dexter M, Siriwardena AN, Read RC. Strategies to increase influenza vaccination rates: outcomes of a nationwide cross-sectional survey of UK general practice. *BMJ Open*. (2012) 2:e000851. doi: 10.1136/bmjopen-2011-000851
- Miller E, Salisbury D, Ramsay M. Planning, registration, and implementation of an immunisation campaign against meningococcal serogroup C disease in the UK: a success story. *Vaccine*. (2001) 20:S58–67. doi: 10.1016/S0264-410X(01)00299-7
- Buntain CL, McGrath E, Golbeck J, LaFree G. *Comparing Social Media and Traditional Surveys around the Boston Marathon Bombing. #Microposts* (2016).
- Snelson CL. Qualitative and mixed methods social media research. *Int J Qual Methods*. (2016) 15:160940691562457. doi: 10.1177/1609406915624574
- Conway M, Hu M, Chapman WW. Recent advances in using natural language processing to address public health research questions using social

The views expressed are those of the author and not necessarily those of the NHS, the NIHR or the Department of Health.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fdgth.2021.804855/full#supplementary-material>

- media and consumer-generated data. *Yearbook Med Inform*. (2019) 28:208–17. doi: 10.1055/s-0039-1677918
- Statista. *Active Social Media Users in the United Kingdom (UK) 2021*. (2021). Available online at: <https://www.statista.com/statistics/507405/uk-active-social-media-and-mobile-social-media-users/> (accessed September 15, 2021).
- Data Reportal. *Digital 2021: The United Kingdom*. (2021). Available online at: <https://datareportal.com/reports/digital-2021-united-kingdom> (accessed September 15, 2021).
- Banda JM, Tekumalla R, Wang G, Yu J, Liu T, Ding Y, et al. A large-scale COVID-19 Twitter chatter dataset for open scientific research—an international collaboration. *Epidemiologia*. (2021) 2:315–24. doi: 10.3390/epidemiologia2030024
- Medford RJ, Saleh SN, Sumarsono A, Perl TM, Lehmann CU. An “infodemic”: leveraging high-volume twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak. *Open Forum Infect Dis*. (2020) 7:ofaa258. doi: 10.1093/ofid/ofaa258
- Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS ONE*. (2010) 5:e14118. doi: 10.1371/journal.pone.0014118
- Kim EH-J, Jeong YK, Kim Y, Kang KY, Song M. Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *J Inform Sci*. (2016) 42:763–81. doi: 10.1177/0165551515608733
- Velasco E, Agheneza T, Denecke K, Kirchner G, Eckmanns T. Social media and internet-based data in global systems for public health surveillance: a systematic review. *Milbank Q*. (2014) 92:7–33. doi: 10.1111/1468-0009.12038
- Statista. *Countries With Most Twitter Users*. (2021). Available online at: <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/> (accessed September 15, 2021).
- Liu S, Liu J. Understanding behavioral intentions toward COVID-19 vaccines: theory-based content analysis of tweets. *J Med Internet Res*. (2021) 23:e28118. doi: 10.2196/28118
- Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. *J Med Internet Res*. (2020) 22:e19016. doi: 10.2196/19016
- Ariana Rimmel. *Communicating COVID Vaccine Safety Poses a Unique Challenge*. (2021). Available online at: <https://media.nature.com/original/magazine-assets/d41586-021-01257-8/d41586-021-01257-8.pdf> (accessed August 14, 2021).
- Keeping London Safe. *COVID-19 Insight Toolkit: Keep London Safe Campaign*. (2019). Available online at: https://drive.google.com/file/d/11T-6QtGJUio19ir3mabAR_yDII5NPghp/edit (accessed August 14, 2021).
- Twitter. *Twitter API*. (2021). Available online at: <https://developer.twitter.com/en/docs/twitter-api> (accessed October 07, 2021).
- Wordnerds. *Meaningful Insights for the Best Customer Experience*. (2021). Available online at: <https://www.wordnerds.ai/> (accessed September 15, 2021).
- Wordnerds. *Big Six Energy CX During COVID-19 Summary Report: A Twitter Study Using Deep Learning, NLP & Linguistics*. (2020). Available online at: <https://www.wordnerds.ai/hubfs/Marketing%20Assets/Wordnerds%20Energy%20Brand%20Report%20-%20June%202020.pdf?hsCtaTracking=2b57c6ee-25fe-410f-a2ee-a8d9f6c8f43d%7Caedfd3b5-f141-404a-b25f-775f73eaf228> (accessed December 22, 2021).
- World Health Organisation (WHO) *SAGE Working Group on Vaccine Hesitancy. Report of the SAGE Working Group on Vaccine Hesitancy* (2014).

32. Freeman D, Loe BS, Chadwick A, Vaccari C, Waite F, Rosebrock L, et al. COVID-19 vaccine hesitancy in the UK: the Oxford coronavirus explanations, attitudes, and narratives survey (Oceans) II. *Psychol Med.* (2020). doi: 10.1017/S0033291720005188. [Epub ahead of print].
33. Local Government Association. *Confidence, Complacency, Convenience Model of Vaccine Hesitancy.* (2021). Available online at: <https://www.local.gov.uk/our-support/coronavirus-information-councils/covid-19-service-information/covid-19-vaccinations/behavioural-insights/resources/3Cmodel-vaccine-hesitancy> (accessed August 14, 2021).
34. Curtis HJ, Inglesby P, Morton CE, Mackenna B, Walker AJ, Morley J, et al. Trends and clinical characteristics of COVID-19 vaccine recipients: a federated analysis of 57.9 million patients' primary care records in situ using OpenSAFELY. *medRxiv [Preprint]*. doi: 10.1101/2021.01.25.21250356
35. Royal Society for Public Health Vision VaPR. *New Poll Finds Ethnic Minority Groups Less Likely to Want COVID Vaccine.* (2020). Available online at: <https://www.rsph.org.uk/about-us/news/new-poll-finds-bame-groups-less-likely-to-want-covid-vaccine.html> (accessed August 12, 2021).
36. Abram MD, Mancini KT, Parker RD. Methods to integrate natural language processing into qualitative research. *Int J Qualit Methods.* (2020) 19:160940692098460. doi: 10.1177/1609406920984608
37. Ghosh D, Guha R. What are we 'tweeting' about obesity? Mapping tweets with topic modeling and Geographic Information System. *Cartogr Geogr Inform Sci.* (2013) 40:90–102. doi: 10.1080/15230406.2013.776210
38. Tavoschi L, Quattrone F, D'Andrea E, Ducange P, Vabanesi M, Marcelloni F, et al. Twitter as a sentinel tool to monitor public opinion on vaccination: an opinion mining analysis from September 2016 to August 2017 in Italy. *Hum Vaccines Immunother.* (2020) 16:1062–9. doi: 10.1080/21645515.2020.1714311
39. Mueller A. *WordCloud for Python Documentation.* (2020). Available online at: https://amueller.github.io/word_cloud/index.html# (accessed December 14, 2021).
40. Raghupathi V, Ren J, Raghupathi W. Studying public perception about vaccination: a sentiment analysis of tweets. *Int J Environ Res Public Health.* (2020) 17:3464. doi: 10.3390/ijerph17103464
41. Piedrahita-Valdés H, Piedrahita-Castillo D, Bermejo-Higuera J, Guillem-Saiz P, Bermejo-Higuera JR, Guillem-Saiz J, et al. Vaccine hesitancy on social media: sentiment analysis from June 2011 to April 2019. *Vaccines.* (2021) 9:28. doi: 10.3390/vaccines9010028
42. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci.* (2011) 6:42. doi: 10.1186/1748-5908-6-42
43. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the theoretical domains framework of behaviour change to investigate implementation problems. *Implement Sci.* (2017) 12:77. doi: 10.1186/s13012-017-0605-9

Author Disclaimer: The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

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.

Copyright © 2022 Lanyi, Green, Craig and Marshall. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.