

Computational Model of Engagement With Stigmatised Sentiment

COVID and General Vaccine Discourse on Social Media

Straton, Nadiya

Document Version Final published version

Published in: Network Modeling Analysis in Health Informatics and Bioinformatics

DOI: 10.1007/s13721-024-00456-3

Publication date: 2024

License CC BY

Citation for published version (APA): Straton, N. (2024). Computational Model of Engagement With Stigmatised Sentiment: COVID and General Vaccine Discourse on Social Media. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 13(1), Article 21. https://doi.org/10.1007/s13721-024-00456-3

Link to publication in CBS Research Portal

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 03. Jul. 2025









ORIGINAL ARTICLE



Computational model of engagement with stigmatised sentiment: COVID and general vaccine discourse on social media

Nadiya Straton¹

Received: 10 July 2023 / Revised: 3 February 2024 / Accepted: 4 March 2024 © The Author(s) 2024

Abstract

The growth rate of new social media users continues to surpass new Internet users and new unique mobile phone subscribers and this trend remains consistent over the past 5 years (2019–2023). The most frequently visited types of websites or apps worldwide are chat and messaging, closely followed by social networks and this trend has also remained relatively constant. The dominating role of social media, especially as a source for information seeking, is staggering, particularly during the COVID-19 pandemic. However, the research in Keller et al. (J Mec Internet Res 16:e8, 2014) indicates that not many experts consider social media as a tool for sharing their expertise or for integrating social media into their research efforts. This is a troubling fact, especially considering that stigmatised health narrative are fueled in the face of uncertainty and spread very quickly among the lay population. The latter contributes to the spread of misinformation and, consequently, fosters hesitancy about preventive measures such as vaccines. This research presents new evidence on engagement with stigmatised vaccine discourse on Facebook (Meta), Twitter (X), YouTube and Reddit. Engagement with health-related sentiment can be an important indicator of perceptions regarding preventive measures. The current research can draw the attention of public health experts to the connection between stigmatised discourse and engagement in health discussions, as well as the potential impact of other linguistic features on engagement. It can also guide health authorities and medical professionals in developing effective communication strategies for the general public. Moreover, there are only a handful of studies discussing engagement with socially shared health-care discussions. The research focuses on examining engagement with stigmatised sentiment in vaccination discussions during and prior to the COVID-19 pandemic, using a cross-sectional approach. The study is based on primary data from social media domains, supplemented by secondary data analyses of literature related to the topic. To achieve the research goal, the study employs a multi-method design primarily based on quantitative methods of content analyses, such as Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al. in Linguistic inquiry and word count: Liwc 2001, vol 71. Lawrence Erlbaum Associates, Mahway, 2001) to extract general language features and analyse stigmatised sentiment as the result of previous research findings in Straton et al. (Stigma annotation scheme and stigmatized language detection in health-care discussions on social media. In: Proceedings of The 12th Language Resources and Evaluation Conference (LREC, 2020), pp 1178–1190 (European Language Resources Association, 2020), Straton (Appl Intell, pp 1–26, 2022). Unsupervised K-means clustering methods, along with F-score and Z-score measures, are applied to draw insights from the features. The goal is to explore the phenomena of engagement with socially shared health information, investigate whether stigma can affect the engagement, and identify any other language features that may have an impact. Highly engaging general vaccine discussions before the pandemic appear to be more planned and less emotional. They exhibit reasoning and differentiation typical of more complex discussion sections, incorporating references to authority and family concerns simultaneously. The engaging messages convey stigmatised sentiment, likely shared either with the intent to deceive, or are based on a genuinely biased belief. Some elements of stigmatised discussions indicate deception, while others are more frequent in truthful statements. Engaging content in COVID vaccine discussions, similar to general vaccine content, lacks emotional elements and exhibits high linguistic complexity. However, there seems to be no connection between engagement and stigmatised sentiment. This difference is likely due to the removal of public anti-vaccination pages during the pandemic where anti-vaccination groups moved to discussion section of posts that try to disprove vaccine conspiracy theories. Mixing polarised groups will result in different feedback and engagement with the content. Almost half of the features in COVID and general vaccine discussions have a similar impact on engagement. However, further exploration of the findings suggests

Extended author information available on the last page of the article

that there are distinct differences. Understanding the differences is very important to prevent generalisation errors in the conclusions drawn from the data. Engagement with written text on social media should be measured within a narrow scope: limited regional, demographic and temporal scope. Otherwise, interpretation of the findings risks on being inconsistent.

Keywords Engagement \cdot Stigma \cdot Facebook (Meta) \cdot Twitter (X) \cdot Reddit \cdot YouTube \cdot Sentiment analyses \cdot Regression \cdot K-means

1 Introduction

Over the past 5 years, the number of social media users has increased by more than a billion, which is greater than the increase in new Internet users and unique mobile phone subscribers over the same period, as seen in Fig. 1.

YouTube, Facebook (Meta), Twitter (X), and Reddit are ranked among the ten most visited websites based on website traffic, according to the Semrush's ranking as of April 2023 (Simon Kemp, 2023). Over the past 5 years, from 2019 to 2023, these platforms consistently maintain their positions in the top ten, with some small variations. Notably, Twitter (X) did not appear in the list in 2020 while Reddit was not among the top ten in 2019. The total monthly average visits by Internet users are as follows: YouTube (2nd place with 94.8 billion visits after Google), Facebook (Meta) (3rd place with 13.8 billion visits), Twitter (X) (6th place with 8.52 billion visits) and Reddit (9th place with 5.41 billion visits) (Simon Kemp, 2023).

Over the past 4 years, Facebook (Meta) has been the world's most used social media platform followed by You-Tube (Simon Kemp, 2023). Facebook's monthly active users account for 37.2% share of all people on Earth, totalling 2.989 billion active users. However, this share increases if users under thirteen are not accounted for (due to their restricted access) and users from China (where Meta is blocked). The top five countries with the most Facebook users, according to data from April 2023, are India (370 million), the United States (186 million), Indonesia (135

	Å Å&			8 8 8 A
	TOTAL POPULATION	UNIQUE MOBILE PHONE SUBSCRIBERS	INDIVIDUALS USING THE INTERNET	SOCIAL MEDIA USERS
2019 October	7.73 billion +1.0% Fr. Oct 18 +79 million	5.15 billion +2.4% fr. Oct 18 +123 million	4.48 billion +10.0% Fr. Oct 18 +416 million	3.72 billion +9.6% Fr. Oct 18 +328 million
2020 October	7.81 billion +1.0% Fr. Oct 19 +81 million	5.20 billion +2.0% 1 Fr. Oct 19 +102 million	4.66 billion +7.4% ↑ Fr. Oct 19 +321 million	4.14 billion +12.3% Fr. Oct 19 +453 million
2021 October	7.89 billion +1.0% Tr. Oct 20 +80 million	5.29 billion +1.9% Tr. Oct 20 +97 million	4.88 billion +4.8% Tr. Oct 20 +222 million	4.55 billion +9.9% Tr. Oct 20 +409 million
2022 October	7.99 billion +0.8% 1 Fr. Oct 21 +66 million	5.48 billion +3.2% fr. Oct 21 +170 million	5.07 billion +3.5% Tr. Oct 21 +171 million	4.74 billion +4.2% Fr. Oct 21 +190 million
2023 October	8.06 billion +0.9% Fr. Oct 22 +72 million	5.60 billion +2.7% Fr. Oct 22 +145 million	5.30 billion +3.7% Tr. Oct 22 +189 million	4.95 billion +4.5% Fr. Oct 22 +215 million

Fig. 1 Growth and digital development trends 2019–2023: world population development, unique mobile phone subscribers, internet usage, social media users. Infographic is based on information collected from Global Digital Reports series, from DataReportal by Simon Kemp (2023)

million), Brazil (114 million) and Mexico (93 million) (Simon Kemp, 2023). Similarly, the share of Facebook audience is the highest in the respective regions, with 22.1% in Southern Asia, 18.5% in South-Eastern Asia, 11.6% in Southern America, 9.3% in North America and 5.4% in Central America (Simon Kemp, 2023). The share of male users on Facebook is 56.8%, while female users account for 43.2%, which is a more balanced distribution than prevailing male audiences on Reddit and Twitter (X) platforms. The median age of Facebook users is 32 years, with the largest share belonging to the 25-34 age group constitutes the largest share (29.6%), followed by the 18-24 age group (22.6%), and the 35–44 age group (19.0%) (Simon Kemp, 2023). While there is a considerable share of young people using Facebook, it is mostly the older population that considers Facebook their favourite platform; the share of older audience is also growing. Engagement by a 'typical' user on Facebook is eleven "likes", five "comments", and one "share" per month. Females users tend to be more active on social media, engaging by liking more content and writing more comments, according to Meta advertising resources data from July 2021 (Simon Kemp, 2023).

Country of origin determines level of engagement to a greater extent than gender. In Greece, a single post received an average of sixteen likes in 2021, while in South Korea—only two likes (Simon Kemp, 2023). People from the Faroe Islands appeared to be more active, liking twenty-four posts per month (median), whereas people in Japan liked only two posts, according to data from June 2021. Since vaccine data in the current study is in English, it will most likely be responded to by an audience from the United States, one of the countries with the most registered users on the platform with an average number of ten likes per post, per month for a 'typical' user, according to a July 2021 report.

When comparing the average number of likes on the platform, both pro-vaccine and anti-vaccine discussions on Facebook (Meta) might fall below the engagement level of a 'typical' user from the United States. Notably, pro-vaccine discourse tends to attract higher engagement, especially if the sentiment contains stigma (Table 1). However, the main interest of current study is engagement with laymen discussions (comments), rather than the frequency of liking by a 'typical' user. Another aspect is the difference in average engagement per post and comment data. Upon observation, it is noted that the engagement with a comment is usually lower than with a post, Table 1 illustrates that the average engagement varies with 9 likes for pro-vaccine and 4 likes for anti-vaccine discussions.

The Reddit platform is the 9th most visited website according to the latest Semrush's ranking as of December 2023, with 7.23 billion visits (Semrush, 2023). On average, users spend just under eight minutes on the site per visit (Mercado, 2023). Most of the Reddit users come from United States (48.69%) followed by the UK (7.05%), Canada (6.99%), Australia (4.19%) and Germany (3.14%) according to Skillademia data from March 2023 (Mercado, 2023) with 52 million daily active users (Simon Kemp, 2023). In the current research, Reddit is the most male dominated social media platform with 65.02% male and 34.98% female users (Mercado, 2023). The largest age group on Reddit are users aged 18 to 29, followed by 30 to 49 range (22% of the traffic) (Lin, 2023).

Roughly 4.6% of all people on Earth use Twitter (X), the share will be higher if not to account for users younger than thirteen (where access to the platform is limited) (Simon Kemp, 2023). Most users that use Twitter (X) are from the United States (64.9 million), followed by Japan (51.8 million), Brazil (16.6 million), and India (15 million) (Simon Kemp, 2023). The Twitter (X) platform is predominantly used by male users (64.3%), while female users make up more than one-third (35.7%) (Simon Kemp, 2023).

YouTube is the second most popular social media platform in the world according to Q4, 2023 data (Statista, 2023) with roughly 31,5% of people on Earth using You-Tube (Simon Kemp, 2023). Most users of YouTube coming from the same top five countries as users of Facebook (Meta): India, the United States, Brazil, Indonesia and Mexico). 54.4% of YouTube's users are male and 45.6% are female (based on April 2023 data) (Simon Kemp, 2023). Another report based on survey reveals that 48.6% of YouTube viewers are men and 51.4% are women. Users aged 25–34 constitute 20.7% of YouTube's audience, followed by 35–44 (16.7%), and 18–24 (15.0%).

The current research does not track, age, gender and place of origin of the social media audience. However, the

Table 1 Engagement with stigmatised sentiment, general vaccination discourse

Dataset Description		Total Data		Stigma		Not Stigma		Undefined		
Facebook (Meta)	#Comments	Min	Max	Average	Engagement Range	Average	Engagement Range	Average	Engagement Range	Average
Anti-Vaccine Discourse	1,399	0	160	4	0–160	5.90	0-41	1.41	0–63	1.74
Pro-Vaccine Discourse	1,264	0	349	9	0–349	11.34	0–218	6.09	0-83	2.45
Total	2,663	0	349	6	0–349	8.43	0–218	3.43	0-83	2.06

Dataset description		Total L	Data		Stigma		Not Stigma		Undefined	
Response to COVID Vaccine Stigma	#Comments	Min	Max	Average	Engagement Range	Average	Engagement Range	Average	Engagement Range	Average
Reddit	4,814	0	2,576	7.95	0-1751	8.42	0-2576	7.91	0 - 278	5.72
Fwitter (X)	2,252	0	630	5.04	0-138	3.46	0-617	5.42	0-630	5.25
YouTube	1,788	0	751	12.85	0-684	12.65	0-751	13.59	0-256	6.41
Total	8,854	0	2,576	8.20	0-1751	8.13	0-2,576	8.40	0-630	5.77
Response to disproving COVID vaccine stigma	#Comments	Min	Max	Average	Engagement range	Average	Engagement range	Average	Engagement range	Average
YouTube	31,230	0	10,390	25.42	0-10,205	21.81	0-10,390	26.37	0-9182	26.92
Fotal	40,084	0	10,390	21.62	0-10,205	18,68	0-10,390	22.47	0-9182	21.82

🖄 Springer

discussions are conducted in English and will attract users who consume content in the language. Since the United States and India have large user base across all discussed platforms one can expect users from those countries to have an impact on the engagement figures in the current research. Gender distribution on social media platforms might also be reflected in engagement figures (Table 2). YouTube, with the highest share of female users, who are also more actively engaged in sharing and liking the content, might explain the much higher engagement rate in Table 2. The largest share on Facebook (Meta), Reddit, Twitter (X), and YouTube is the young audience in the age range up to 35 years. One of the top three main reasons for individuals in the age range from 24 to 35 years to use social media is to read news.

To a large extent, Facebook (Meta) users post and share photos/videos with friends and family and then keep up to date with the news on the platform according to data from October, 2023 (Simon Kemp, 2023). Audience on Reddit looks for funny/entertaining content and subsequently keeps up to date with the news (Simon Kemp, 2023), whereas Twitter (X) is used primarily to keep up to date with the news (Simon Kemp, 2023). Most YouTube subscribers gather information, knowledge on the platform, with one-third using platform to watch news (Mahajan, 2023).

As social media channels in the current research are primarily used for receiving information and news from social media platforms, there should be a level of trust in the information received, since most likely, their decisions will be based on it.

Therefore, it is quite important to study engagement with health information shared on social media platforms, especially regarding vaccines, as it can impact health-care choices, such as the decision to vaccinate and subsequently determine the future course of a pandemic. Literature review showed that only a small group of studies looked into engagement with health content on social media. One third of researchers do not consider information from social media to be of value for their research efforts (Keller et al., 2014). Even though researchers might recognise the potential of social media data, a majority is either sceptical or oppose obtaining information from social media platforms (Keller et al., 2014). In the interim, when researchers are not professionally engaged in the social media space, public discourse is shaped by various laymen opinions regarding important health-related issues, which are often treated as factual information.

1.1 Social media and engagement with health information

Before the widespread use of social media channels, patients increasingly turned to the internet and reported that healthrelated information they found online was as useful, if not more so, than advice received from their doctors (Keller et al., 2014). One-way communication from the official channels to the public, coupled with a lack of discussions about the content, might contribute to a greater divide between patient seeking information on social media platforms and health-care experts who provide the content without prior insight into the patient's perspective.

As previously discussed, the largest demographic using social media platforms fall within the age range of 18–35, with one-third of this group utilising social media to seek content and read news/stories (Simon Kemp, 2023). There are reports that young people demonstrate greater engagement with health-care content on social media due to improved learning, socialisation, and emotional support, and that such information impacts them even if they quickly scroll through it (Goodyear et al., 2018). Given that official organisations were reported to have the most impact on the behaviour of young people (Goodyear et al., 2018), one needs to examine communication strategies between official channels and the public, leading to greater participation of health professionals in social media discussions. The authors in Wong et al. (2014) discuss collaboration with health professionals and patients as a potential remedy against misinformation on social media. They suggest looking for cues beyond social media platforms such as health rating sites, to understand what different age groups expect from health-care providers. If authors in Wong et al. (2014) report that a better understanding of the patient's perspective on health is needed, the authors in Keller et al. (2014) note that there is very little evidence on how public health professionals use social media and communicate information with patients. All known studies on social media engagement with health information have reported the need for a better understanding of the relationship between different agents and health content. Keller et al. (2014), Wong et al. (2014), Goodyear et al. (2018), Pérez-Escoda et al. (2020) discuss the need to understand the relationship between patients, health content and official agents on social media better. Straton et al. (2017a, b, c) discuss engagement of people with health information based on big data from Facebook (Meta), however these studies only look into characteristics of the posted content rather than sentiment analyses.

The ongoing research aims to bridge the gap in comprehension by exploring the content and its attributes that effectively captivate the public within health-care conversations. It's objective is to provide guidance to health experts on the optimal methods for communicating with the general public. The type of content, and especially the meaning assigned to it (Miller et al., 2016), can have an overwhelming impact. The current study attempts to explore whether prejudiced sentiment about vaccines can impact engagement with the content. Additionally, it seeks to explore other features in a text that can capture attention and trigger a response. Health experts on platforms like YouTube, for example, attempt to disprove prejudiced beliefs about COVID vaccines. However, the information is often presented through the one-way communication in videos, where official channels aim to disprove conspiracies by addressing prejudices about vaccines in general and COVID vaccines in particular. Discussions among laypeople are often left with no further involvement from public health professionals or official channels. Various quantitative methodologies are being employed in the present study to ascertain the efficacy of textual discourse on social media platforms. This investigation is anticipated to yield insights that will enhance our understanding of communication strategies directed at the general public, thereby facilitating adherence to evidence-based health-care practices and informed decision-making regarding personal health and preventive measures such as vaccines.

Based on the author's comprehensive understanding, the primary contributions of the study can be summarized as follows:

- 1. This is the first study on the engagement with healthcare content studied on a number of social media platforms with different length and structure of the comments: Facebook (Meta), Reddit, Twitter (X), YouTube.
- 2. The first research that studies engagement with stigmatised sentiment and other textual features in the context of general vaccine and COVID vaccines discussions in particular.
- 3. The study provides a meaningful insight on engagement with socially shared health information and its variance with time, discusses demographics and why it is important to study the concept within a narrow context.
- 4. The findings are supported with a publicly shared dataset.

The first contribution is partially addressed in Sect. 2 the Dataset, 3 Methods, and 4 Discussion sections. The second contribution is discussed in Sects. 3, 4: Methods and Discussion section. The third contribution is addressed in Sects. 4, 5: Discussion, Conclusion, and partially in Abstract, 1 Introduction and Sect. 1.1. The forth contribution is discussed in Sect. 2 the Dataset.

2 The dataset

The data were collected before the COVID-19 pandemic from the largest Facebook (Meta) pages discussing general vaccines. One Facebook (Meta) wall exhibited an anti-vaccine stance, while another presented a pro-vaccine sentiment, as shown in Table 1. The data were collected from January 2018 to February 2019.

Despite the public nature of the pages, they exhibit strong in-group support, with anti-vaccine and pro-vaccine

sentiment prevailing on their respective pages. This in-group support and alignment of views lead to homogeneous opinions, seldom welcoming open discussions or opposing views on the topic. Notably, stigmatised content triggers responses and results in higher engagement on average, particularly evident in stigmatised anti-vaccine discussions.

The COVID vaccine data outlined in Table 2 were collected after the outbreak of the COVID-19 pandemic and subsequent endeavors aimed at eliminating anti-vaccination pages from social media platforms. The data spans the period from April 2020 to March 2021, covering roughly nine months before the COVID vaccine roll-out and three months after. Out of the 117 main pages on Facebook (Meta) discussing vaccines, 26 were removed (Straton, 2022), and some pages also changed their view settings from public to private. As a result of de-platforming the most influential pages, the number of followers of anti-vaccine pages decreased on average, while the number of followers on provaccination pages increased (Straton, 2022).

After the removal of the most influential pages from Facebook (Meta), followers of anti-vaccination pages shifted their discussions to express their stance on pro-vaccination pages (Straton, 2022). This behaviour is not typical for the anti-vaccine movement, as observed in general vaccine discussions before the COVID pandemic outbreak. The theories of cognitive dissonance (Festinger, 1962) and selective exposure (Freedman and Sears, 1965) suggest that dealing with contrasting ideas can be emotionally and psychologically exhausting. As a result, the logical reaction is either to avoid circumstances with contradictory information or to make a decision not to question prior beliefs when confronted with opposing views (Festinger, 1962). Confirmation bias, which is part of cognitive dissonance theory, demonstrates the tendency to interpret information that supports one's own beliefs while ignoring alternatives. Group members also tend to believe they have no bias and perceive themselves as morally superior to the out-group (Festinger, 1962).

On the other hand, Abrams and other authors in Abrams et al. (1990), Abrams and Hogg (2010) have discussed the positive social identity linked to self-esteem, which is consequently enhanced through a sense of identification with a specific group and 'intergroup' discrimination. Therefore, the sense of belonging to an anti-vaccine group and the experience of cognitive dissonance when confronted with opposing views can fuel conflict and discrimination, hindering constructive discussion between the groups. Allport, in his contact hypothesis, identified four preconditions that should be met in order to facilitate inter-group understanding: equal status of the groups in a given situation, common goals, absence of intergroup competition, and support from authorities or social norms (Allport et al., 1954). The first three preconditions can be challenging to achieve in the context of vaccine stigma. Elliot Aronson added two additional preconditions related to frequent interaction between the in-group and outgroup, as well as mutual interdependence (Aronson et al., 1994). When popular anti-vaccine pages were removed from social media, the movement became somewhat dependent on pro-vaccine pages, resulting in frequent interaction between the vaccine groups. However, anti-vaccine and pro-vaccine discussions did not show signs of intergroup cooperation during the pandemic. Authorities and society were interested in limiting anti-vaccine bias and erroneous information about the vaccines, making the possibility of equal status unrealistic. Even if the authorities did not impose limits on the antivaccine presence, cooperation between the groups is highly unlikely as the anti-vaccine and pro-vaccine groups pursue different goals even during the pandemic, which is one of the necessary preconditions for finding common ground.

During the COVID-19 pandemic, social media users responded to vaccine information by liking, disliking, or re-sharing the content, discussing events in the comments. Engagement levels with 'not stigma' and 'stigma' content are about the same with small variations (Table 2). Response to disproving COVID conspiracy sentiment shows slightly higher engagement with 'not stigma' sentiment (Table 2).

Engagement behaviour within general vaccine discourse shows a distinctly higher trend of engagement with stigmatised sentiment among those with an anti-vaccine stance (Table 1). Before the pandemic, both anti-vaccine and provaccine groups primarily discussed vaccines within their respective groups, with minimal to no participation from the out-of-group. In such environments, strong biases towards vaccines in anti-vaccine groups could be expressed among like-minded group members without much reservation. This environment promotes confirmation bias and allows the avoidance of alternative points of view and contradictory ideas.

The higher engagement with stigmatised sentiment among anti-vaccine group members could be attributed to the homogeneous nature of discussions, reinforcing their existing beliefs about vaccines and views targeted against out-of-group members, such as governments and pharmaceutical companies. Hence higher endorsement of stigmatised sentiment. Another reason could be the different nature of social media structure, even though, all information analysed in the research is publicly accessible.

Within Facebook (Meta), a convenient mechanism exists for facilitating discussions within its own in-group, whether public or private. This platform features a singular route for incoming and outgoing traffic, along with restrictions on the number of friends permitted to follow an account. The environment is characterised by a symmetric structure, which may foster heightened in-group engagement compared to alternative social media platforms such as Twitter (X). According to Paul and Friginal (2019), discussions on Facebook (Meta) are observed to be more interactive than those on Twitter (X). This observation is predominantly corroborated in relation to stigmatised sentiment, as indicated in Tables 1 and 2.

On Twitter (X), information is received asymmetrically through various channels (Peters et al., 2013), which may create opportunities for inter-group interaction. Additionally, there are no constraints on the number of followers a particular account may have, and there is no obligation to reciprocally follow those who follow the account.

During the COVID-19 pandemic, there was increased interaction between anti-vaccine and pro-vaccine groups, often responding to the same content. Whether engaging with posts perpetuating stigma surrounding COVID vaccines or those debunking conspiracy theories related to COVID vaccines, a prevalent anti-vaccination stance and distinct arguments in favour of vaccinations are frequently observed.

When the pro-vaccine movement becomes involved in discussions, reactions to stigmatised anti-vaccine sentiment become diverse. This diversity may account for the relatively consistent engagement levels, on average, across stigma, non-stigma, and undefined sentiment in response to COVID vaccine posts. However, there is a slightly elevated level of engagement observed with non-stigmatised sentiment, particularly on Twitter (X) and YouTube platforms (Table 2). Further investigation is conducted using ANOVA F-score regression to determine if stigmatised sentiment can be a good predictor of engagement, if the difference in engagement with non-stigma sentiment is significant for COVID vaccine data, and if there are noticeable variations in the features across different levels of engagement prior to and during the COVID-19 pandemic.

3 Methods

3.1 Main language features with LIWC

Engagement concept holds significant importance in this study, as it serves as an indicator of participation in vaccine discussions and elucidates the factors that contribute to increased involvement in online debates. Diverse attributes are gathered from multiple social media platforms to measure engagement, as specified in Table 3. The engagement attribute in the propagated dataset is derived from upvotes/ downvotes (Reddit), likes (Twitter (X), YouTube), dislikes (YouTube), and retweets (Twitter (X)).

The views feature on social media platforms, such as YouTube, is not directly related to engagement with the content, as it is unclear whether the person viewed the content intentionally or opened it by mistake. Therefore, it was disregarded in the study. Negative engagement values are converted into positive values due to the recognition that negative engagement, like its positive counterpart, signifies a response to the content.

The engagement feature is log-normalised to eliminate skewness from the highly variable data and is presented as different engagement levels for better visualisation of the results. The engagement levels are established based on the distribution of values in the dataset. Their levels differ slightly in general anti-vaccination discussions (low engagement: $0 \le and \le 2$; medium engagement: $2 \le and \le 4$; high engagement: $4 \le 1$ and pro-vaccination discussions (low engagement: $0 \le$ and < 2.5; medium engagement: 2.5 <= and < 5; high engagement: 5 <=). Anti-COVID vaccine discussions with following engagement levels (low engagement: $0 \le and \le 2$; medium engagement: $2 \le and \le 5$; high engagement: $5 \le 3$ and discussions that aim to disprove COVID anti-vaccine sentiment (low engagement: $0 \le and$ < 3; medium engagement: 3 <= and < 6; high engagement: 6 <=).

 Table 3 Engagement measures on various social media domains

Facebook (Meta)	Number of Likes per Comment	Engagement range per sentiment	Average number of reactions per sentiment	Engagement range per Facebook page	Average number of reactions per Facebook page
Reddit	Number of Upvotes and Downvotes per Comment	Engagement range per sentiment	Average number of reactions per senti- ment	Engagement range per Reddit page	Average number of reactions per Reddit page
Twitter (X)	Number of Likes and Retweets per Com- ment	Engagement range per sentiment	Average number of reactions per senti- ment	Engagement range per dataset	Average number of reactions per dataset
YouTube	Number of Likes and Dislikes per Com- ment	Engagement range per sentiment	Average number of reactions per senti- ment	Engagement range per dataset	Average number of reactions per dataset

	Ratio of variance (btw. Engagement val. and Features)	Z-score: Standard Error of the MEAN				
Features	Anova F-score regression	High Eng.	Med. Eng.	Low Eng.		
Word Count	88.6	6.9342	5.4961	-4.8020		
No. characters	86.19	6.7149	5.4950	-4.7351		
No. of Words per Sentence	57.5	8.3158	6.0989	-5.5278		
Stigma, Not Stigma, Und. sentiment	36.06	-0.5340	1.1103	16.7038		
Prepositions (with, above)	27.1	3.5939	3.7365	-2.9314		
Function Words (of, and)	19.46	4.1537	3.9004	-3.1820		
Affect as Process (ugly, bitter)	13.67	-3.3200	-3.6921	2.8264		
Family references	8.84	2.3369	2.0483	-1.7182		
Clout (power)	6.35	2.4020	1.7634	-1.5976		
Punctuation mark (the period)	5.77	-2.4466	-1.3918	1.4280		
Health references (flu, cough)	4.89	2.1814	2.6991	-1.9918		
Article	4.7	3.3895	1.6307	-1.8317		
Positive Emotions (happy, good)	4.57	-1.9720	-2.3298	1.7462		
They pronoun	4.56	3.0651	2.8682	-2.3431		
% words captured by the dictionary	4.55	1.9519	1.6291	-1.3949		
Assent (agree, yes)	4.25	-1.8957	-2.5153	1.8145		
Adjectives (free, long)	4.14	-1.5400	-1.7532	1.3310		
Exclamations	4.04	-1.5735	-2.5106	1.7145		
Personal Pronoun (them, itself)	3.92	1.6656	1.1357	-1.0649		
Netspeak (lol, thx)	3.79	-1.7954	-1.7082	1.3863		
Differentiation (hasn't, else)	3.72	1.5487	0.9144	-0.9203		
Female references (girl, her)	2.67	-0.3613	1.4675	-0.6136		
Space (down, in)	2.42	0.9811	2.5804	-1.5692		
Pronouns	2.34	0.8802	0.8637	-0.6926		
Conjunctions (but, whereas)	2.29	1.8180	2.4015	-1.7349		
AllPunc	2.21	-1.6358	-0.5816	0.7828		
Discrepancy (should, could)	2.18	0.9916	0.5659	-0.5797		
Perceptual Processes (touch, listen)	2.08	0.4527	3.0910	-1.6605		
Swear words (damn, shit)	1.84	-1.6353	-1.9154	1.4399		
Second-person pronoun you	1.65	0.5116	0.3670	-0.3360		

In order to understand which features are highly dependent on the response variable, or in other words which features are relevant for the analyses of engagement ANOVA F-score (F-value) is calculated using a slightly simplified version from Ding et al. (2014), Kumar et al. (2015):

$$F = \frac{SSA/df_b}{SSW/df_w} = \frac{\sum_{j=1}^k n_j (\bar{X}_j - \bar{X})^2 / df_b}{\sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2 / df_w} = \frac{MSA}{MSW}$$
(1)

There are several features with data points. Dispersion between data points is established with sum of squares. *SSA*—sum of squares among groups (features), *SSW*—sum of squares within groups (features), *MSA*—mean of squares among, *MSW*—mean of squares within, *df*—degrees of freedom, $(\bar{X}_i - \bar{X})$ —distance between each feature average value

 X_j and grand means, X_{ij} —distance between each observed value within the feature from the feature mean \bar{X}_j .

If H_0 is true, variances are equal and if there is equal variance between feature and response variable, it means that feature has no impact on the engagement and can be disregarded from the model. The ANOVA F-ratio will be close to 1. However, if $P \leq 0.05$, the null hypothesis is rejected and relevant features selected. Implementation is performed in Python with SelectKBest and f_classif functions imported from feature_selection module from sklearn package. Out of the 84 features in Table 6, 30 features were selected (Tables 4, 5).

While ANOVA F-value measures dependency between different features and continuous engagement variable. Z-score provides an insight on the type of dependency at different engagement levels, where engagement is converted into categorical variable (high, medium, low) and Z-score is calculated based on Wang and Chen (2012):

$$Z = (x - \mu)/\sigma$$

where x is an individual value, μ is population mean, σ is population standard deviation. Before the COVID-19 pandemic, messages with stigmatised sentiment generated higher engagement compared to messages without stigma, particularly stigmatised anti-vaccine discussions, which achieved higher engagement (HighEng. -0.5340; MedEng. 1.1103; LowEng. 16.7038 as presented in Table 4). Higher Z-score levels indicate lower stigma in the text. The ratio of variance (ANOVA F-score regression) between engagement values and stigmatised sentiment demonstrates that stigma can be a relevant feature and predictor of engagement (Table 4) based on data from general vaccine discussions. Stigmatised sentiment is identified as part of semiautomated process of manual annotation in a smaller subset of data and propagation of the labels to a larger social media dataset with supervised learning techniques as described in Straton (2022). The annotated and propagated data are stored in the Figshare depository (Straton 2023).

Greedy k-means++ algorithm helps to visualise connection between engagement attribute and various features not only limited to stigmatised sentiment in a text. It is an unsupervised way to show hidden structure in the input data and distribution of data points into clusters, where one can possibly see connection between the features if any (Bhattacharya et al., 2019). With each iteration greedy k-means++ algorithm samples several new centers, and then greedily chooses the one that decreases the objective the most (Arthur et al., 2007):



1: Uniformly independently sample $c_1^1, ..., c_1^l 1 \in X$;

2: Let $c_1 = argmin_{c_{\tau}\{c_1,...,c_{t}\}}\varphi(X, c)$ and set $C_1 = \{c_1\}$.

3: for $i \leftarrow 1, 2, 3, ..., k - 1$ do

4: Sample $c_i^1 + 1, ..., c_i^l + 1 \in X$ independently, sampling *x* with probability $\varphi(x, C_i)/\varphi(X, C_i)$;

5: Let $c_i^1 + 1 = argmin_{c_{\in}\{c_i^1, \dots, c_i^l\}} \varphi(X, C_i \cup \{c\})$ and set $C_{i+1} = C_i \cup \{c_{i+1}\}.$

6: return $C := C_k$ (Grunau et al. 2023)

l candidate centers are sampled $c_{i+1}^1, ..., c_{i+1}^l$ from the constructed distribution in every step. For each candidate center c_{i+1}^j the new cost of the solution $\varphi(X, C \cup \{c_{i+1}^j\})$ is computed. Subsequently, the candidate center that minimises the solution is selected (Grunau et al., 2023; Arthur et al., 2007).

The greedy k-means++ algorithm is implemented using the KMeans function imported from the cluster module, in Scikit-learn library (Pedregosa et al., 2011) in Python. Four clusters, ten initialisation, and a maximum three hundred iterations are utilised. The figures in 2D (Figs. 2, 3) depict the relationship between engagement, stigma, and one additional feature.

Figure 2a, b show that highly engaging messages contain a higher number of function words. In contrast, messages with more positive emotions do not elicit the same level of response, as messages without. Highly engaging messages show low level of positive emotions. This trend is consistent with the patterns observed in COVID vaccine discussions.

Some features that predict engagement with COVID vaccines can also predict engagement with general vaccines; however, there are also differences. Stigmatised sentiment does not affect engagement in COVID vaccine discussions, whereas it is associated with higher engagement in general



(a) Engagement with Function Words and Annotated Stigma Feature



(b) Engagement with Positive Emotion and Annotated Stigma Feature

Fig. 2 General vaccine data, prior to the pandemic

7 coores Standard Error of the MEAN

Datio of variance (htw

Table 5COVID vaccines(data from Twitter (X),YouTube, Reddit 2020–2021):top 30 features that predictEngagement

Footures

reatures	Engagement val. and Features)					
	ANOVA F-score regression	High Eng.	Med. Eng.	Low Eng.		
Article	10380	6.2121	5.8002	-4.3784		
You, second-person pronoun	5160	-5.2487	-3.5367	3.0530		
Comma	2370	2.1106	4.0313	-2.4642		
Biological Processes	2260	0.7311	-0.3792	-0.0121		
Positive emotions (happy, good)	2150	-3.8795	-7.0482	4.3579		
Number	2090	-2.7721	-1.1546	1.2744		
Function Words (of, and)	1840	9.0892	9.8624	-7.0583		
Time (hour, day)	1830	3.6215	4.4357	-3.0522		
No. words per sentence	1830	6.5468	10.9510	-6.9072		
I, first-person singular	1810	5.9202	1.6052	-2.3139		
Affect as process (ugly, bitter)	1800	-4.8818	-7.0271	4.6109		
Money (audit, cash)	1520	0.4743	1.1524	-0.6705		
Insight (know, consider)	1460	2.2412	2.0217	-1.5461		
Health references (clinic, flu)	1430	0.9465	0.5528	-0.5103		
Causation (because, effect)	1420	-0.0549	3.0453	-1.4288		
Clout (power words)	1380	3.3647	0.1924	-0.9739		
She, he pronouns	1330	8.4436	-2.2125	-1.1667		
Assent (agree, yes)	1260	-2.4068	-3.3206	2.2050		
Sentiment score (polarity)	1140	2.5684	-2.4525	0.4884		
Home (house, kitchen)	1050	1.9085	-1.1710	0.0542		
Hearing (listen, sound)	1040	1.9730	-4.1633	1.4554		
Sexuality (love, incest)	1020	0.4936	0.3395	-0.2904		
Male references (boy, his)	1020	11.0819	-3.6343	-1.1850		
Focuspresent (today, now)	940	2.1052	1.4515	-1.2401		
Netspeak (lol, thx)	870	-4.2454	-5.5824	3.7592		
Drives	850	-0.3265	2.1276	-0.9226		
Exclamation	780	-1.8804	-2.6741	1.7606		
Preposition (with, above)	760	6.6381	9.7061	-6.3412		
Tentative (uncertain, not definite)	700	1.7759	5.7253	-3.1791		
Conjunctions (but, whereas)	650	1.5040	7.1508	-3.7834		

vaccine discussions. Stigmatised sentiment prevailed in conversations within both anti-vaccine and pro-vaccine groups prior to the COVID-19 pandemic (Fig. 2). Anti-vaccine movement openly expressed stigmatised views regarding vaccines and the out-group without encountering opposing opinions in the discussion stream.

In Straton (2022), it is observed that stigmatised sentiment does not influence engagement with the content. This is evidenced by the absence of significant variances in engagement across stigma and non-stigma annotation labels, as indicated by *z*-score and ANOVA F-score. Users exhibited similar reactions to discussions on COVID vaccines that contained high levels of stigma and those that were free from prejudice towards vaccines.

While COVID vaccine discussions void of stigma exhibit slightly higher engagement, z-score for

🖄 Springer

different engagement levels is not significant, as evidenced in Table 6 (HighEng. 2.0312; MedEng. -0.3827; LowEng. -0.3515).

Furthermore, the ratio of variance (ANOVA F-score regression) of 270 between engagement values and the stigma, not stigma features is relatively insignificant compared to the other top 30 features in Table 5. The stigma feature is not a strong predictor of engagement based on COVID vaccine discussion data and is one of the least correlated features with the engagement.

This is further supported by the results of visualisation using unsupervised learning K-means clustering on the engagement feature, annotated class, and one of the top features that predict engagement. The visualisation examples in Fig. 3a, b depict positive emotion and function words features.

Table 6	Features of the model	and their performance	within each engagement	cluster (Z-score, Star	ndard Error of the MEAN)	

		Z-score: CO	OVID Vaccine		Z-score: Ge	eneral Vaccines	
Туре	Features	High Eng.	Medium Eng.	Low Eng.	High Eng.	Medium Eng.	Low Eng.
Other Features	Stigma, Not Stigma, Und. Sentiment	2.03	-0.38	-0.35	-0.53	1.11	16.70
Other Features	No. characters	-2.45	5.65	-2.03	6.71	5.50	-4.74
Other Features	Sentiment Score (polarity)	2.57	-2.45	0.49	1.20	0.60	-0.66
Other Features	Subjectivity Score	4.66	4.17	-3.20	-1.30	-1.93	1.34
Summary Dimensions	Word Count	-2.42	5.30	-1.88	6.93	5.50	-4.80
Summary Dimensions	Analytical Thinking	-2.01	-0.89	0.95	0.18	-1.43	0.65
Summary Dimensions	Clout: power	3.36	0.19	-0.97	2.40	1.76	-1.60
Summary Dimensions	Authentic	1.12	4.29	-2.33	1.59	1.25	-1.10
Summary Dimensions	Emotional Tone	2.85	-0.70	-0.42	-0.16	-0.98	0.53
Summary Dimensions	Words per Sentence	6.55	10.95	-6.91	8.32	6.10	-5.53
Summary Dimensions	Words > 6 Letters	-0.11	2.32	-1.07	-0.46	0.12	0.08
Summary Dimensions	% words (dictionary)	8.35	5.37	-4.73	1.95	1.63	-1.39
Linguistic Processes	Function words	9.09	9.86	-7.06	4.15	3.90	-3.18
Linguistic Processes	Pronoun	6.19	2.08	-2.61	0.88	0.86	-0.69
Linguistic Processes	Ppron (them, itself)	4.90	0.08	-1.32	1.67	1.14	-1.06
Linguistic Processes	I	5.92	1.61	-2.31	0.27	-0.27	0.05
Linguistic Processes	We	1.17	2.77	-1.62	0.86	-0.39	-0.07
Linguistic Processes	You	-5.25	-3.54	3.05	0.51	0.37	-0.34
Linguistic Processes	SheHe	8.44	-2.21	-1.17	-0.57	0.13	0.11
Linguistic Processes	They	-0.32	4.48	-2.04	3.07	2.87	-2.34
Linguistic Processes	Ipron (me, my)	3.37	2.86	-2.24	-0.62	-0.05	0.21
Linguistic Processes	Article	6.21	5.80	-4.39	3.39	1.63	-1.83
Linguistic Processes	Prepositions (with, above)	6.69	9.71	-6.34	3.59	3.74	-2.93
Linguistic Processes	Auxverb (may, must)	1.56	4.59	-2.58	1.18	1.87	-1.28
Linguistic Processes	Adverbs (really, qucikly)	3.96	2.05	-2.01	-1.21	-0.40	4.15
Linguistic Processes	Conjunctions (but, whereas)	1.50	7.15	-3.78	1.82	2.40	-1.73
Linguistic Processes	Negate (not. never)	-1.95	-1.60	1.27	-0.08	-0.41	0.22
Other Grammar	Verbs	3.54	3.14	-2.42	1.11	1.60	-1.13
Other Grammar	Adjectives (free, long)	-0.99	0.22	0.16	-1.54	-1.75	1.33
Other Grammar	Number	-2.77	-1.15	1.27	-0.03	-0.73	0.37
Other Grammar	Quantifiers (few many)	-0.16	2.45	-1.12	1.17	1.86	-1.27
Affect	Affective Processes (ugly, bitter)	-4.88	-7.03	4.61	-3.32	-3.69	2.83
Affect	Positive Emotions (happy, good)	-3.88	-7.05	4.36	-1.97	-2.33	1.75
Affect	Negative Emotions (resent, enemy)	-2.73	-2.26	1.79	-1.17	-0.47	0.58
Affect	Anxiety (afraid, tense)	-0.29	0.36	-0.10	-0.95	1.17	-0.29
Affect	Anger (rage, hurt)	-2.04	-2.65	1.79	-0.09	0.61	-0.27
Affect	Sadness (grief, crv)	0.61	0.15	-0.23	-0.60	-1.16	0.75
Social	Social Processes (talk, friend)	6.16	-2.06	-0.64	1.01	0.87	-0.74
Social	Family	-0.41	-0.48	0.34	2.34	2.05	-1.72
Social	Friend	7.61	-3.05	-0.55	-0.29	-1.24	0.70
Social	Female references (girl, her)	-1.29	-1.75	1.16	-0.36	1.47	-0.61
Social	Male references (boy, his)	11.08	-3.63	-1.18	-0.71	-0.64	0.53
Cognitive Processes	Cognitive Processes (cause, ought)	2.89	6.26	-3.73	1.06	0.29	-0.47
Cognitive Processes	Insight (know, consider)	2.32	2.02	-1.55	0.66	-0.37	-0.02
Cognitive Processes	Causation (because, effect)	-0.06	3.05	-1.43	0.24	2.06	-1.09
Cognitive Processes	Discrepancy (should could)	1.92	2 18	-1.54	0.99	0.57	-0.58
Cognitive Processes	Tentative (nerhaps guess)	1.72	5 73	-3.18	1.06	-0.43	-0.11
Cognitive Processes	Differentiation (hasn't, else)	-0.39	3.19	-1.41	1.55	0.91	-0.92

		Z-score: CC	OVID Vaccine		Z-score: Ge	neral Vaccines		
Туре	Features	High Eng.	Medium Eng.	Low Eng.	High Eng.	Medium Eng.	Low Eng.	
Perceptual Processes	Perceptual Processes (touch, listen)	2.43	-4.97	1.72	0.45	3.09	-1.66	
Perceptual Processes	Seeing (view, look)	2.26	-2.38	0.53	-1.10	1.21	-0.27	
Perceptual Processes	Hearing (listen, sound)	1.97	-4.16	1.45	-0.71	-0.90	0.66	
Perceptual Processes	Feeling (hold, felt)	0.39	-0.95	0.35	-0.11	-0.06	0.06	
		Z-score: CC	OVID Vaccine		Z-score: Ge	Z-score: General Vaccines		
Туре	Features	High Eng.	Medium Eng.	Low Eng.	High Eng.	Medium Eng.	Low Eng.	
Biological Processes	Health (clinic, flu)	0.95	0.55	-0.51	2.18	2.70	-1.99	
Biological Processes	Sexuality (love, incest)	0.49	0.34	-0.29	-0.43	0.42	-0.07	
Drives	Drives	-0.33	2.13	-0.92	0.59	0.73	-0.54	
Drives	Affiliation (ally, friend)	0.98	0.84	-0.66	0.78	-0.18	-0.15	
Drives	Achievement (win, success)	1.10	0.39	-0.47	0.27	-0.27	0.05	
Drives	Power (superior, bully)	-0.98	1.38	-0.40	0.12	1.90	-0.97	
Drives	Reward (prize, benefit)	-0.49	-0.91	0.56	-0.59	0.71	-0.17	
Drives	Risk (danger, doubt)	-0.03	3.23	-1.52	0.38	0.66	-0.44	
Time Orientation	Focuspast (ago, did)	2.89	1.06	-1.26	-0.10	0.46	-0.19	
Time Orientation	Focuspresent (today, now)	2.11	1.45	-1.24	-0.16	2.86	-1.36	
Time Orientation	Focusfuture (will, soon)	-1.47	1.39	-0.27	1.40	-0.76	-0.05	
Relativity	Motion (walk, move)	0.28	1.35	-0.71	0.83	2.85	-1.66	
Relativity	Space (down, in)	0.61	2.26	-1.23	0.98	2.58	-1.57	
Relativity	Time (hour, day)	3.62	4.43	-3.05	0.58	0.15	-0.25	
Personal Concerns	Work (work, boss)	0.16	2.99	-1.46	0.84	1.22	-0.85	
Personal Concerns	Leisure (house, music)	1.08	-2.78	1.03	-0.26	0.55	-0.19	
Personal Concerns	Home (house, kitchen)	1.91	-1.17	0.05	0.43	1.51	-0.88	
Personal Concerns	Money (audit, cash)	0.47	1.15	-0.67	0.53	1.53	-0.91	
Personal Concerns	Religion (altar, church)	-0.93	-3.64	1.97	-0.84	-1.32	0.91	
Personal Concerns	Death (bury, kill)	-1.25	-1.40	0.99	0.81	0.61	-0.54	
Filler words	Swear words (damn, shit)	-2.50	-2.58	1.88	-1.64	-1.92	1.44	
Filler words	Netspeak (lol, thx)	-4.24	-5.58	3.76	-1.80	-1.71	1.39	
Filler words	Assent (agree, yes)	-2.41	-3.32	2.20	-1.90	-2.52	1.81	
Filler words	Nonfluencies (uh, rr*)	-1.62	-3.52	2.09	-0.80	-1.19	0.83	
Filler words	Fillers (blah, you know)	-0.94	-1.36	0.89	-0.61	-1.25	0.80	
Punctuation Marks	AllPunc	-1.82	-3.11	1.95	-1.64	-0.58	0.78	
Punctuation Marks	Period	-1.93	-2.51	1.69	-2.45	-1.39	1.43	
Punctuation Marks	Comma	2.11	4.03	-2.46	0.25	-0.19	0.02	
Punctuation Marks	QMark	-2.78	-3.31	2.30	-1.01	-1.67	1.13	
Punctuation Marks	Exclam	-1.88	-2.67	1.76	-1.57	-2.51	1.71	
Punctuation Marks	Apostro	4.31	5.49	-3.73	-0.76	0.94	-0.23	
Punctuation Marks	OtherP	-2.49	-1.30	1.27	-0.40	1.76	-0.75	

 Table 6 (continued)

Users react similarly to stigmatised and non-stigmatised sentiment, as observed in Fig. 3a, b. Across different engagement values and engagement clusters, the ratio of stigma, non-stigma, and undefined sentiment remains largely consistent, as indicated by Z-scores, ANOVA F-score regression, and unsupervised learning K-means clustering. However, messages with a high ratio of function words typically demonstrate medium to high engagement. Conversely, discussions characterised by a high ratio of positive emotions tend to garner lower engagement with the content.



(a) Engagement with Function Words and Annotated Stigma Feature



(b) Engagement with Positive Emotion and Annotated Stigma Feature

Fig. 3 COVID vaccine data

4 Discussion

Data regarding general vaccines have been collected from Facebook (Meta) and displayed in Table 1, whereas comments related to COVID vaccines Straton (2023) were gathered from Reddit, YouTube, and Twitter (X) (Table 2).

Different engagement levels in Tables 4 and 5 are presented for better visualisation and are derived from log-normalised engagement attribute. Yet, the relationship between features and engagement values is discerned through regression analysis. Engagement with stigmatised content, especially on antivaccination pages, was quite high before the pandemic. The anti-vaccination and pro-vaccination movements predominantly shared information within their own groups, rarely interacting with out-of-groups in order to avoid cognitive dissonance. Amidst the COVID-19 pandemic, the elimination of anti-vaccination pages (Straton, 2022) resulted in increased interactions with pro-vaccine groups in the comment sections. This transition might have influenced engagement with vaccine stigma content.

Throughout the COVID-19 pandemic, there is consistently higher engagement with not stigmatised sentiment on Twitter (X), and YouTube platforms in Table 2. The Z-score standard error of the mean shows slightly higher engagement with not stigma sentiment than stigmatised sentiment based on COVID vaccine data in Table 6 (a lower Z-score indicates stigmatised sentiment, while a higher Z-score indicates not stigmatised sentiment).

Nonetheless, an ANOVA F-score regression of 270 indicates that the variance between the stigma feature and engagement is insignificant compared to other features in COVID vaccine discussions (Table 5), suggesting that

the stigma feature is not a particularly strong predictor of engagement.

The significance of the stigmatised feature as a predictor of engagement in general vaccine discussions before the pandemic, juxtaposed with its diminished relevance after the COVID-19 pandemic, suggests that the concept of engagement should be studied within a specific context. Efforts have been undertaken previously to analyse the interaction with health-related content on social media platforms. These studies have utilised large dataset extracted from Facebook (Meta) in order to discern the potential attributes of comments and posts that contribute to increased engagement. Straton et al. (2017a, 2017b) identified the primary factors influencing engagement as post type, a time span between post creation and post update, and a time of a day the post was created. Post types on Facebook (Meta) included short status updates, other textual type posts, photos, videos, and link types posts. Although visual content was identified as the most engaging in the study (Straton et al., 2017b), highly engaging textual messages, such as short texts, were also observed. The discoveries from Straton et al. (2017c) indicate that the sentiment conveyed in social media messages can influence engagement.

Thus, when delving deeper into textual features, it becomes evident that certain language features contribute more to engagement than others. Nevertheless, these features vary in vaccine discussions before and after the pandemic, which could be attributed to specific events. The removal of anti-vaccine pages from social media platforms during the COVID-19 pandemic resulted in both anti-vaccination and pro-vaccination debates occurring within the same comment threads of COVID vaccine-related posts. The latter certainly changed the engagement dynamics with vaccine content on social media. If stigmatised language against out-of-group is encouraged before the pandemic through in-group discussions, mixed in-group and out-of-group debates during the COVID pandemic no longer show any engagement variance for this type of sentiment. Various factors may contribute to alterations in engagement dynamics, whereby features of the text preceding and following the pandemic may offer additional context into these fluctuations.

ANOVA F-score regression analyses was utilised to identify the most relevant features impacting engagement, with the Z-score (representing the standard error of the mean) employed as the measure of engagement variation. Engagement levels were subsequently categorised as high, medium, and low, predicated upon log-normalised engagement values and dataset features.

Prior to the pandemic, twelve features emerged as robust predictors of engagement (Table 4): number of words per sentence, word count, number of characters in the message, stigma/not stigma sentiment, prepositions (with, above), function words (of, and), affect as process (ugly, bitter), family references, clout (power), health references (flu, cough), article, and the pronoun 'they'.

During the COVID-19 pandemic, the features demonstrating the greatest influence on engagement, as determined by ANOVA F-score regression and Z-score analyses (Table 5), encompass: article, punctuation mark (comma), pronouns 'she' and 'he', function words (of, and), time (hour, day), number of words per sentence, first-person pronoun 'I', prepositions (with, above), conjunctions (but, whereas), positive emotions (happy, good), affect as process (ugly, bitter), and second-person pronoun 'you'.

There is an overlap among some of the top thirty features in both datasets, yet their significance, as assessed through ANOVA F-score regressions, demonstrates variations across the datasets. These common features include: number of words per sentence, prepositions (with, above), function words (of, and), affect as process (ugly, bitter), clout (power words), health references (clinic, flu), article, positive emotions (happy, good), assent (agree, yes), exclamations, personal pronouns (them, itself), netspeak (lol, thx), and conjunctions (but, whereas).

Even though, 43% of the features exhibit significant correlations in both datasets, the majority of features do not overlap and hold predictive capacity concerning engagement within the specific context of discussions concerning COVID vaccines or general vaccines. This observation suggests that a nuanced understanding and an effective measurement of engagement with content on social media platforms are best attained by considering specific features of the text or when exploring engagement within a specific context. It aligns with previous findings in Straton et al. (2017b) and Straton et al. (2017c), indicating that textual features can influence engagement. Exploring engagement through the analysis of big data offers broad insights into the phenomenon; the multitude of factors influencing engagement can yield varied outcomes contingent upon shifts in the contextual framework under study.

Features prevalent in general vaccine data before the pandemic (Table 4) suggest that high engagement is associated with an increased number of words per sentence, as well as a higher number of words and characters count in a message. The latter can be linked to verbal fluency, cognitive complexity, or dominance in the conversation (Tausczik and Pennebaker, 2010).

In a cooperative context, a higher word count feature indicates better coordination in communication. However, in a discussion context, a higher word count may suggest a breakdown in the discussion (Tausczik and Pennebaker, 2010). Vaccine discussions resemble conversations with elements of debate and diatribe, suggesting a need for lengthier argumentation to convince and persuade.

High engagement in general vaccine discussions is strongly linked to prejudice, stereotypes, and stigmatised messages, eliciting responses from users. Such behaviour can be expected in highly disputed topics, especially considering that the discussions primarily took place within in-group settings. Prepositions also play a prominent role in highly engaging comments. According to Hartley et al. (2003) in Tausczik and Pennebaker (2010), prepositions signify complex and detailed information, typically found in the reasoning and discussion parts of written text. The prevalence of function words in highly engaging comments indicates how individuals communicate and allocate their attention, as reflected in their use of pronouns. The frequent use of 'they' pronouns in engaging texts suggests an awareness of the out-group. As mentioned in Gunsch et al. (2000) and cited in Tausczik and Pennebaker (2010), an increase in references to the other party (in this instance, 'they') suggests a concentration on an adversary. In diatribes, this focus is likely to be negative. Family references can be attributed to discussions about parents' vaccine choices for their children, while clout/power words may suggest control and the authority of government institutions in administering vaccines. Health references, such as 'clinic' and 'flu', are present in engaging comments in both general and COVID vaccine discussions, and they are also found in low engagement discussions. Health references are commonly observed in discussions about health and vaccines. The prevalence of articles in highly engaging posts suggests a tendency to be very specific through the utilisation of concrete nouns, as discussed in Pennebaker and Lay (2002) and Tausczik and Pennebaker (2010).

According to Newman et al. (2003) and Pennebaker and Lay (2002), articles, pronouns, and prepositions can reveal important information about a person, similar to relevant nouns or verbs. The observation that the use of articles and prepositions is linked to fewer emotion words, as discussed in Tausczik and Pennebaker (2010), finds support in the current research. Highly engaging posts contain fewer positive and other emotion words, including swear words, which are considered negative words according to Rassin and Muris (2005) in Oliver et al. (2008). Affective processes are associated with emotionality, as discussed in Blonder et al. (2005), Djikic et al. (2006), and Gill et al. (2008), with citation in Tausczik and Pennebaker (2010). The findings in Table 4 indicate a negative correlation between highly engaging discussions and the presence of affect words and emotions. This suggests that highly engaging posts are less likely to be written impulsively and are written with more self-control.

Highly engaging general vaccine discussions appear to be more planned and less emotional, featuring reasoning and differentiation typical of more complex discussion sections, also incorporating simultaneous references to authority and family concerns. The engaging sentiment aligns with the stigmatised sentiment described in Straton (2022). Stigma can be intentionally shared to deceive, or the person may genuinely believe in biased information being shared.

Engaging content in COVID vaccine discussions, similar to general vaccine discourse, often lacks emotional elements. This is evident based on the negative correlation observed between positive emotion features, affective processes, and engagement categories. Moreover, the prevalence of articles and prepositions in the text suggests the use of less emotional language, as discussed in Tausczik and Pennebaker (2010). The presence of articles may also indicate higher linguistic complexity. The number of words per sentence feature, also observed in general vaccine discussions, implies that more words are employed to persuade, as vaccine discussions often resemble diatribes. However, as discussed in Hancock et al. (2007), this could also indicate deception when comparing the message characteristics of liars and truth tellers. Contrary to expectations, the use of the pronoun 'you' in the current research is linked with low engagement, as indicated by ANOVA F-score regression in Table 5. As discussed in Tausczik and Pennebaker (2010), pronouns can signal the quality of the relationship. Excessive use of the pronoun 'you' can indicate blaming or distancing attitudes, as mentioned in Hahlweg et al. (1984) and cited in Simmons et al. (2005). Simmons et al. (2005) elaborated that the 'you' pronoun suggests negative interaction behaviour. The finding could elucidate the lower engagement observed in discussions characterised by in-group and out-group dynamics, when messages contain a predominance of second-person pronouns.

The use of the first-person singular pronoun 'I' is correlated with high engagement and, as mentioned in Tausczik and Pennebaker (2010), it is more commonly used in discussions of lower status. Additionally, the first-person singular pronoun is also associated with honesty; less frequent use of the first-person singular pronoun is associated with deception, according to Newman et al. (2003) and later confirmed by Hancock et al. (2007). Since the firstperson pronoun signifies 'taking ownership of the statement', liars tend to avoid using it in a discussion (Hancock et al. 2007).

There is a positive correlation between high engagement and the use of she/he pronouns. According to Bond and Lee (2005), third-person pronouns (he, she, her) are less common in deceptive statements and more frequent in truthful messages.

Messages that exhibit high levels of engagement also feature an increased frequency of references to temporal indicators. According to the findings in Vrij et al. (2007) and Vrij (2005), contextual references, including mentions of time, appear more frequently in truthful discussions than in deceptive ones.

Conjunctions are pervasive in engaging comments and play a role in logically organising words, thereby contributing to the coherence of the overall statement, as observed in Graesser et al. (2004), Tausczik and Pennebaker (2010). Additionally, highly engaging comments tend to have more punctuation marks, such as commas, which can be attributed to the relatively longer sentences used in these discussions.

Emotionally neutral content tends to be highly engaging when discussing COVID vaccines. While a high number of words might suggest deception, other features, such as references to time, the use of first-person singular pronouns, and the use of third-person singular pronouns, indicate honesty in the discussion. Blaming and distancing attitudes, often conveyed through the pronoun 'you', are infrequently encountered in comments that exhibit medium to highl levels of engagement. Moreover, highly engaging COVID-19 vaccine discussions exhibit coherence and include features such as the use of articles, which may indicate the complexity of the language (Pennebaker and King, 1999).

The extent to which the results of the analyses encompass demographics remains uncertain, given the absence of demographic data collection in the current study. As social media channels mature, the user cohort using them also ages. However, despite of extensive range of demographics already using social media platforms, there are many groups who do not have access to it. Primarily, young adults and middle age groups are receiving their health-information from social media domains and it is possible that analyses is disregarding older age groups from the study. Furthermore, the data sourced from Facebook (Meta), Reddit, and Twitter (X) might exhibit 'gender-bias' owing to the predominance of male audiences on these platforms.

5 Conclusion

Previous studies (Keller et al., 2014; Wong et al., 2014; Goodyear et al., 2018; Pérez-Escoda et al., 2020) advocate for increased involvement in examining interactions with health-care discourse, researcher engagement, youth participation, and regional interest in health-care content. Additionally, several quantitative inquiries have explored engagement with health information disseminated on social media (Straton et al., 2017a, b, c). Nevertheless, these subsequent studies have not delved into the sentiment of the discourse. The COVID-19 pandemic emphasised the pivotal significance of socially shared health discourse, particularly due to the prevalent use of social media for obtaining health-related information. This study not only reaffirms the significance of previous research but also sheds light on how discussions among the general public rapidly morph into perceived expert viewpoints, possibly influenced by the relatively low involvement of the research community. Filtering out pages containing highly stigmatised opinions about health can help foster a more balanced perspective; however, it may not completely eradicate the issue. Stigmatised narratives about health on social media are fuelled by uncertainty and dichotomisation of topics, such as vaccines, especially newly developed ones. Despite the efforts to remove misleading content, it finds its way back.

In order to gain insights into the features that capture the attention of the general public regarding health-care information, as well as to comprehend their reactions to stigmatising narratives, this study utilised a combination of quantitative techniques, such as ANOVA F-score regression, Z-score, and K-means clustering. The integration of all three approaches aids in providing a holistic understanding of engagement with stigmatised sentiment, alongside other sentiment features, within health discussions shared across social media platforms.

Prior to the COVID-19 pandemic, vaccine discussions featuring stigmatised, prejudiced, and stereotyped content often garnered high engagement. Consequently, there was an increase in references to blame, attributable to the in-group nature of such discussions. There was little to no reservation in stigmatising the out-group, since stigmatised discussions are easily echoed among like minded group members. Ingroup communication on social media platforms can serve as echo-chambers, perpetuating fear and consequently creating challenges for health organisations and practitioners in convincing those within the group to make informed decision about their health. High engagement with stigmatised content suggests an imbalance in discussions, indicating a lack of fact-checking with health authorities.

Studying the public's engagement with vaccine discourse or any health-care information is pivotal, as it provides health authorities with valuable insights into the most impactful sentiment features to utilise when communicating with the general public or participating in pertinent discussions. Engaging discussions reference authority and family, whereas stigma is primarily directed against pharmaceutical companies, government institutions, and members of the out-group movement. Such content is more structured, complex, less emotional, and 'polished' compared with low-engaging content. The sentiment features indicate that comments were not spontaneously written, with certain premeditation, which is a significant finding in the study.

Certain events, such as the removal of anti-vaccination pages after the COVID-19 pandemic began, can alter the dynamics of a discussion and the content that is liked and shared. Additionally, stigmatised sentiment, alongside other features, appears to exert a limited influence on engagement in vaccine discourse during the COVID pandemic. The conclusions drawn in the current research underscore the significance of investigating engagement with social media content within a narrowly defined context, given the challenges of accounting for evolving dynamics over time.

Current research serves as a reminder to health authorities about the importance of avoiding stereotyped and prejudiced sentiment when communicating with the public about health care and prevention measures such as vaccines in particular. Persistent polarisation surrounding vaccines fueled by stigma and prejudice, can results in decisions with broader implications beyond individual vaccine hesitancy.

One potential direction for future research is to build upon the findings of this study and investigate engagement with socially shared health discourse within a narrower demographic or regional contexts. This would entail extracting data from social media platforms that are relevant to specific regions and analysing the language features within that context.

Future research on engagement with health-care content should foster more collaboration between computer scientist, social scientists, patients, and health-care providers. This might requite undertaking longitudinal studies, such as monitoring patients' exposure to stigmatised healthrelated content (including vaccines) and analysing their subsequent health-care decisions, taking into account their demographic traits.

Closer collaboration between public health experts, researchers, doctors, and laypeople discussing vaccine or other health-related topics on social media can serve as fact checking mechanism. When expert opinions are part of the discussion, such collaboration can lead to a reduction in stigmatised sentiment and more effective communication. Acknowledgements Not applicable.

Author Contributions The author Nadiya Straton confirms sole responsibility for the following: (1) study conception and design of the article; (2) data collection, analysis and interpretation of results; (3) manuscript drafting and preparation, revising it for important intellectual content; (4) review and approval of the final version; (5) all aspects of the work in ensuring that the questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding Open access funding provided by Copenhagen Business School. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of supporting data Analyses are based on data stored on the Figshare repository: https://doi.org/10.6084/m9.figshare.23277392.

Declaration

Ethical review and approval for the study Ethical review and approval for the study were not deemed necessary because the analyses were conducted on publicly available social media data.

Conflict of interest The author declares no conflict of interest in relation to the current work: no competing financial and/or non-financial interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

- Abrams D, Hogg MA (2010) Social identity and self-categorization. SAGE Handb Prejudice Stereotyp Discrimin 1:179–193
- Abrams D, Hogg MA et al (1990) An introduction to the social identity approach. Soc Ident Theory Constr Crit Adv 1:1–9
- Allport GW, Clark K, Pettigrew T (1954) The nature of prejudice
- Aronson E, Wilson TD, Akert RM (1994) Social psychology: the heart and the mind. HarperCollins College Publishers
- Arthur D, Vassilvitskii S et al (2007) k-means++: the advantages of careful seeding. In: Soda 7:1027–1035
- Bhattacharya A, Eube J, Röglin H, Schmidt M (2019) Noisy, greedy and not so greedy k-means++. arXiv preprint arXiv:1912.00653
- Blonder LX et al (2005) Affective facial and lexical expression in aprosodic versus aphasic stroke patients. J Int Neuropsychol Soc 11:677–685
- Bond GD, Lee AY (2005) Language of lies in prison: linguistic classification of prisoners' truthful and deceptive natural language. Appl Cogn Psychol 19:313–329
- Ding H, Feng P-M, Chen W, Lin H (2014) Identification of bacteriophage virion proteins by the anova feature selection and analysis. Mol BioSyst 10:2229–2235

- Djikic M, Oatley K, Peterson JB (2006) The bitter-sweet labor of emoting: The linguistic comparison of writers and physicists. Creat Res J 18:191–197
- Festinger L (1962) A theory of cognitive dissonance, vol 2. Stanford University Press
- Freedman JL, Sears DO (1965) Selective exposure. In: Advances in experimental social psychology, vol. 2. Elsevier, pp 57–97
- Gill AJ, French RM, Gergle D, Oberlander J (2008) The language of emotion in short blog texts. In: Proceedings of the 2008 ACM Conference on computer supported cooperative work, pp 299–302
- Goodyear VA, Armour KM, Wood H (2018) Young people and their engagement with health-related social media: New perspectives. Sport Educ Soc 24:673–688
- Graesser AC, McNamara DS, Louwerse MM, Cai Z (2004) Cohmetrix: analysis of text on cohesion and language. Behav Res Methods Instrum Comput 36:193–202
- Grunau C, Özüdoğru AA, Rozhoň V, Tětek J (2023) A nearly tight analysis of greedy k-means++. In: Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp 1012–1070 (SIAM)
- Gunsch MA, Brownlow S, Haynes SE, Mabe Z (2000) Differential forms linguistic content of various of political advertising. J Broadcast Electron Media 44:27–42
- Hahlweg K et al (1984) Development and validity of a new system to analyze interpersonal communication: kategoriensystem für partnerschaftliche interaktion. Marital Interact Anal Modif. Guilford Press New York, pp 182–198
- Hancock JT, Curry LE, Goorha S, Woodworth M (2007) On lying and being lied to: a linguistic analysis of deception in computermediated communication. Discourse Process 45:1–23
- Hartley J, Pennebaker JW, Fox C (2003) Abstracts, introductions and discussions: How far do they differ in style? Scientometrics 57:389–398
- Keller B, Labrique A, Jain KM, Pekosz A, Levine O (2014) Mind the gap: social media engagement by public health researchers. J Mec Internet Res 16:e8
- Kumar M, Rath NK, Swain A, Rath SK (2015) Feature selection and classification of microarray data using mapreduce based anova and k-nearest neighbor. Proc Comput Sci 54:301–310
- Lin Y (2023) Reddit demographics, based on marketing charts. https://www.oberlo.com/blog/reddit-statistics. Accessed 12 Dec 2023
- Mahajan V (2023) https://www.notta.ai/en/blog/youtube-statistics/. Accessed 15 Dec 2023
- Mercado A (2023) Reddit statistics from skillademia based on similarwed data. https://www.skillademia.com/statistics/reddit-statistics/. Accessed 12 Dec 2023
- Miller D et al (2016) How the world changed social media. UCL press
- Newman ML, Pennebaker JW, Berry DS, Richards JM (2003) Lying words: predicting deception from linguistic styles. Pers Soc Psychol Bull 29:665–675
- Oliver EJ, Markland D, Hardy J, Petherick CM (2008) The effects of autonomy-supportive versus controlling environments on selftalk. Motiv Emot 32:200–212
- Paul JZ, Friginal E (2019) The effects of symmetric and asymmetric social networks on second language communication. Comput Assist Lang Learn 32:587–618
- Pedregosa F et al (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830
- Pennebaker JW, King LA (1999) Linguistic styles: language use as an individual difference. J Pers Soc Psychol 77:1296
- Pennebaker JW, Lay TC (2002) Language use and personality during crises: Analyses of mayor Rudolph Giuliani's press conferences. J Res Pers 36:271–282

- Pennebaker JW, Francis ME, Booth RJ (2001) Linguistic inquiry and word count: Liwc 2001, vol 71. Lawrence Erlbaum Associates, Mahway
- Pérez-Escoda A, Jiménez-Narros C, Perlado-Lamo-de Espinosa M, Pedrero-Esteban LM (2020) Social networks' engagement during the covid-19 pandemic in Spain: health media vs healthcare professionals. Int J Environ Res Public Health 17:5261
- Peters K, Chen Y, Kaplan AM, Ognibeni B, Pauwels K (2013) Social media metrics-a framework and guidelines for managing social media. J Interact Mark 27:281–298
- Rassin E, Muris P (2005) Why do women swear? an exploration of reasons for and perceived efficacy of swearing in Dutch female students. Perso Individ Differ 38:1669–1674
- Semrush (2023) Most visited websites in the world ranking. https:// www.semrush.com/website/top/. Accessed 7–15 Dec 2023
- Simmons RA, Gordon PC, Chambless DL (2005) Pronouns in marital interaction: What do "you" and "i" say about marital health? Psychol Sci 16:932–936
- Simon Kemp taK (2023) Datareportal. https://datareportal.com. Accessed 10 Nov 2023
- Statista (2023) Most popular social networks worldwide, ranked by number of active users. https://www.statista.com/statistics/ 272014/global-social-networks-ranked-by-number-of-users/. Accessed 2 Jan 2024
- Straton N (2022) Covid vaccine stigma: detecting stigma across social media platforms with computational model based on deep learning. Appl Intell 53(13):16398–16423
- Straton N (2023) Social media corpus: Stigma identification in vaccination discourse (covid-19). https://doi.org/10.6084/m9.figsh are.23277392.v1
- Straton N, Jang H, Ng R (2020) Stigma annotation scheme and stigmatized language detection in health-care discussions on social media. In: Proceedings of The 12th Language Resources and Evaluation Conference (LREC 2020), 1178–1190 (European Language Resources Association)

- Straton N, Mukkamala RR, Vatrapu R (2017a) Big social data analytics for public health: Comparative methods study and performance indicators of health care content on Facebook. In: 2017 IEEE International Conference on Big Data (Big Data), pp 2772–2777 (IEEE)
- Straton N, Mukkamala RR, Vatrapu R (2017b) Big social data analytics for public health: predicting Facebook post performance using artificial neural networks and deep learning. In: 2017 IEEE International Congress on Big Data (BigData Congress), pp 89–96 (IEEE)
- Straton N, Vatrapu R, Mukkamala RR (2017c) Facebook and public health: A study to understand facebook post performance with organizations' strategy. In: 2017 IEEE International Conference on big data (Big Data), pp 3123–3132 (IEEE)
- Tausczik YR, Pennebaker JW (2010) The psychological meaning of words: Liwc and computerised text analysis methods. J Lang Soc Psychol 29:24–54
- Vrij A (2005) Cooperation of liars and truth tellers. Appl Cogn Psychol 19:39–50
- Vrij A, Mann S, Kristen S, Fisher RP (2007) Cues to deception and ability to detect lies as a function of police interview styles. Law Hum Behav 31:499–518
- Wang Y, Chen H-J (2012) Use of percentiles and z-scores in anthropometry. In: Handbook of anthropometry: physical measures of human form in health and disease. Springer, pp 29–48
- Wong CA, Merchant RM, Moreno MA (2014) Using social media to engage adolescents and young adults with their health. Healthcare 2:220–224

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Nadiya Straton¹

- ☑ Nadiya Straton nadiyastratonandersen@gmail.com; ns.digi@cbs.dk
- ¹ Department of Digitization, Copenhagen Business School, 2000 Frederiksberg, Denmark