

The Prevalence of mRNA Related Discussions During the Post-COVID-19 Era

Dimitrios KOKKINAKIS^{a,1}, Bastiaan BRUINSMA^b and Mia-Marie HAMMARLIN^c

^a*University of Gothenburg and the Centre for Ageing & Health, AgeCap, Sweden.*

^b*Chalmers University of Technology, Sweden*

^c*University of Lund, Department of Communication and Media, Sweden*

ORCID ID: Dimitrios Kokkinakis <https://orcid.org/0000-0003-3579-4739>

Abstract. Vaccinations are one of the most significant interventions to public health, but vaccine hesitancy and skepticism are raising serious concerns for a portion of the population in many countries, including Sweden. In this study, we use Swedish social media data and structural topic modeling to automatically identify mRNA-vaccine related discussion themes and gain deeper insights into how people's refusal or acceptance of the mRNA technology affects vaccine uptake. Our point of departure is a scientific study published in February 2022, which seems to once again sparked further suspicion and concern and highlight the necessity to focus on issues about the nature and trustworthiness in vaccine safety. Structural topic modelling is a statistical method that facilitates the study of topic prevalence, temporal topic evolution, and topic correlation automatically. Using such a method, our research goal is to identify the current understanding of the mechanisms on how the public perceives the mRNA vaccine in the light of new experimental findings.

Keywords. vaccine hesitancy, structural topic modeling, Swedish internet forum, natural language processing, Swedish tweets, mRNA vaccines, event detection

1. Introduction

Vaccine hesitancy and skepticism can be triggered by anxiety about possible side effects and concerns related to novel vaccine technologies, such as the messenger RNA (mRNA) which can be used as a reason for not receiving (the COVID-19) vaccine [1]. The University of Lund study: "Intracellular Reverse Transcription of Pfizer BioNTech COVID-19 mRNA Vaccine BNT162b2 In Vitro in Human Liver Cell Line" [2], published on the 25th of February 2022, has been frequently cited since its release, as a confirmation for the reason for vaccine hesitancy, highlighting a potential misconception that the mRNA vaccine alters the human DNA. Furthermore, vaccine skepticism is often taken online on social networking sites. Therefore, our study aims to identify and discuss Swedish mRNA-related social media posts, that are being published during the period after the Lund study. We use structural topic modeling (STM) to explore Swedish tweets and Swedish discussion posts from a popular social media platform (Flashback) about mRNA-related vaccination. Our aim is to give answer to two major questions: *what patterns emerge in Swedish social media as a response to the Lund study and how mRNA*

¹ Corresponding Author: Dimitrios Kokkinakis, Department of Swedish, Multilingualism, Language Technology, Box 200, 405 30 Gothenburg, Sweden. E-mail: dimitrios.kokkinakis@gu.se.

related discussion topics change over time, to better understand the public perceptions, responses and concerns that arise after the Lund research. We start by applying natural language processing methods to pre-process and classify the mRNA-related narratives and then structural topic modeling to uncover the most prevalent discussion topics and their evolution over time in the Swedish context. STM automatically detects latent topics in the dataset which can be used to investigate the nature of these topics reflected in the mRNA discussions [3] by utilizing an exploratory mixed quantitative-qualitative approach using data collected from Swedish social media. As a methodological approach, STM enable us to identify prevalent topics in the data, followed by a qualitative analysis on the most representative words and posts of each topic which provided us with better and more targeted insights into the pros and cons of public perceptions and concerns about the mRNA vaccine. Our results could be useful to public health experts and pro-vaccine organizations to formulate even more effective policies and strategies to reduce anti-vaccine reactions and boost vaccine acceptance.

2. Vaccine Hesitancy and Skepticism

According to the World Health Organization [4] vaccine hesitancy, “the reluctance or refusal to vaccinate despite the availability of vaccines”, was one of the top ten threats to global health even before the COVID-19 pandemic. Although vaccinations are considered as one of the most significant interventions to public health, vaccine hesitancy and resistance creates serious concerns for a significant portion of the population in many countries, including Sweden. Skepticism about vaccine effectiveness, adverse effects, personal beliefs and conspiratory claims as well as exposure to misinformation, plays an important role in decreasing rates of vaccination [5,6]. Vaccine hesitancy discussions are often taken online and for an increasing number of people, the use of such platforms has become a major source for information related to health protection and vaccinations [7]. The availability of massive digital content in e.g., Twitter or Reddit enables researchers to rapidly analyze and monitor large amounts of data, to e.g., identify and better understand the vaccine-deniers’ arguments against vaccinations which in turn, can rapidly be spread as *rumours* to an even wider audience. False and disputed news or misleading information about vaccination keeps emerging and flowing between people in social media [8]. There is a sense of freedom of self-expression in the use of language, indicated by e.g., the magnitude of (negative) ways to refer to vaccine; here are examples from the Swedish data: *fejkvaccin* ‘fake vaccine’, *bluffvaccin* ‘hoax vaccine’, *fuskvaccin* ‘fraud vaccine’, *förtryckarvaccin* ‘oppressor vaccine’ or *försökskaninvaccin* ‘guinea pig vaccine’.

3. Data

Swedish tweets were downloaded from Feb., 10, 2022 (two weeks before the Lund study was published) to Nov., 10, 2022. The tweets were collected with the keywords *m-?RNA*. * (‘?’ the preceding character is optional; ‘.*’: ≥ 0 characters) or the hashtag *#mRNA* and *lang:sv* (Swedish content). The final tweet data set consisted of 1,700 unique tweets from 730 different users. Apart from the previous, we also collected ca 7,600 unique posts from the popular Swedish forum Flashback (<https://www.flashback.org/>), from 18 different discussion threads, all related to COVID-19 and mRNA vaccination.

3.1. Preprocessing and data cleansing

We preprocessed the data using R 4.2.1. For each tweet and for each post, we stored the text and some relevant metadata such as the date of publication. Duplicate data as well as Swedish stop words and numbers were deleted, while the textual content was turned to lowercase. During a normalization process, identified token variants such as ‘mrna vaccin’ and ‘mrnavaccin’ were converted to a single uniform format, here ‘mrna-vaccin’. Furthermore, the dataset was further tokenized (basically separating punctuation and metadata from words). Multiword expressions, statistically significant collocations and phrasal verbs were also recognized, and their contiguous components were joined with an underscore prior to further processing (e.g., *big_pharma*; *in_vitro* or *spruta_in* ‘to inject’). Posts with less than 3 tokens were removed due to a small search volume. For the structural topic modeling we used the R package *stm* (version 1.3.6) [9].

4. Structural Topic Modeling

Latent Dirichlet Allocation (LDA) [10], is a popular topic modeling method that uses the statistical analysis of textual data to identify themes or topics that occur in a document collection. Structural topic model (STM) has emerged as an extension to LDA allowing the integration of covariates into the prior distributions for document-topic compositions and topic-word proportions. Thereby, STM can be used to model how the content of a collection of documents changes as a function of document-level covariates such as day and time, and gain insights and understanding on how topics evolve. Since there is no “correct” solution for determining the optimal number of topics k that should be generated during the model selection process, several diagnostic aspects of the topic modeling were evaluated to decide the number of topics, k , to use. The *stm* package implements several evaluation metrics, such as the spread of *semantic coherence* [11] and *exclusivity*, which both capture what humans qualitatively perceive as good topics [9]. After preprocessing of the data, a document-term matrix was created with 7,600 documents, 18,900 terms (i.e., unique words) and used for modeling, while the best model yielded 14 topics. Figure 1a shows the semantic coherence vs the exclusivity of the models, while figure 1b shows the temporal evolution of two of the identified topics.

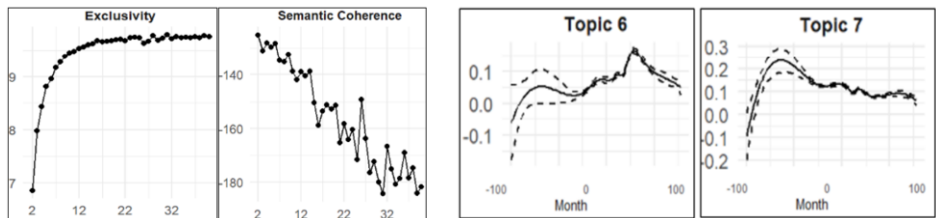


Figure 1a. Semantic coherence and selectivity; figure 1b. Prevalence of topics 6 and 7.

5. Results and Discussion

To summarize and better understand the public responses and concerns that arose after the Lund study, we took a closer look at the results of the STM, in which several general

themes were revealed based on the topics identified. With respect to our first research question, that is *what patterns emerge in Swedish social media as a response to the Lund study*, the most prevalent theme during the first couple of months after its publication, was, as expected, direct related to the Lund article:

- *Oroväckande resultaten från svenska studien: Pfizer-vaccin tar sig in i leverceller - och omvandlas till DNA* ‘Alarming results from the Swedish study: Pfizer vaccine enters livercells - and is converted into DNA’.

A major concern was also the future unknown effects of the mRNA vaccine:

- *Risken är stor att barnen får svåra skador som kan påverka framtida generationer.* ‘The risk is high that children will suffer serious injuries that could affect future generations.’.

Related to the previous is the evidence of the conformity with people’s willingness to get vaccinated, but not with the mRNA vaccine, but rather with a “conventional” one:

- *Jag är absolut ingen vaccinmotståndare, så länge de är traditionellt tillverkade på ett avdödat virus. Men dessa mRNA vacciner är jag väldigt skeptisk till. Tar dom helt enkelt inte!* ‘I am absolutely not against vaccines, as long as they are conventional made from a dead virus. But I am very skeptical about these mRNA vaccines. Just don't take them!’.

Concerns were raised for possible injuries and/or side effects, e.g., on the male genitalia; on the female ovaries and for the traces of mRNA vaccine in breast milk – these were major themes for the period of August-November 2022, due to new research studies:

- *Biverkningar av mRNA-vaccin inkluderar allvarliga skador på penis* ‘Side effects of mRNA vaccines include serious damage to the penis’.

Finally, myocarditis risks were prominent during almost the whole examined period:

- *Läkare varnar för att mRNA vaccin orsakar myokardit men Twitter stämplar inlägget som falsk information.* ‘Doctors warn that mRNA vaccine causes myocarditis but Twitter labels the post as false information.’.

Table 1 shows examples of two of the most prevalent topics with their top-weighted word content as well as an example of a “document”/post, for these topics, which address: the mRNA-to-DNA study; and mRNA’s concerns and risks.

Table 1. Examples of top-rated STM topics

Topic#	Theme	Top words	Top-weighted documents
7	mRNA -to- DNA study	<i>spikeprotein, dna, studie, cell, levercell, omvandla, producera, virus, ta_sig_in</i>	<i>Och vem betalar till människor som fått skador</i> 🚫 <i>Pfizer Vaccine Becomes DNA in Liver Cells.</i> ‘And who pays for people who suffered injuries’ 🚫 <i>Pfizer Vaccine Becomes DNA in Liver Cells.</i> <i>frågan är OM några av oss kommer att kunna återställa det nu när 80% av jordens befolkning även har accepterat förändring av sitt DNA</i> ‘the question is IF some of us will be able to restore it now that 80% of the earth's population has also accepted changing their DNA’
6	Concer ns and risks	<i>myokardit, män, hjärtmuskelinfl., novavax, skada, penis, antivax, studie, foliehatt</i>	<i>Covid-19-vaccinets mRNA i bröstmjölk</i> ‘The covid-19 vaccine’s mRNA in breast milk’ <i>Ja vaccinet samlas i äggstockarna så alla kvinnor bör undvika sex i minst 10 år.</i> ‘Ya the vaccine accumulates in the ovaries so all women should avoid sex for at least 10 years.’

With respect on *how discussion topics change over time* some topics showed some clear characteristics on the topic prevalence (cf. Fig 1b). The time covariate provides a means for a direct comparison and explanatory power over time which makes it better understand people’s concerns during a time span. Retrospectively, the variation of topics across time, can also be helpful to detect significant events related to e.g., mRNA discussions in social media data over time, as in our case, e.g., through an examination of prominent words within each topic. For instance, figure 1b shows how the prevalence of topic 6 raises towards the end of the examined period, while topic 7 rapidly declines

during the middle of the examined period; this is most probably due to the fact that topic 6 relates to the discussion of the potential side effects and risks (e.g., traces of mRNA in breast milk); while topic 7 directly relates or refers to the University of Lund study.

6. Conclusions

This study identifies dominant topics about mRNA vaccine-related issues discussed on Swedish social media. It also examines the changes in these topics over time to better understand the larger trends. Among the selected topics, certain themes, e.g., on myocarditis, remained constant over time. As vaccine development progressed however, other topics became more dominant, driven by e.g., scientific studies introduced to the public. A limitation of the presented work is the search itself which only used a non-exhaustive list of keywords; several relevant posts are probably not included. Consequently, we could expect discussions to be different when using a different keyword set. As a future task, we plan to integrate sentiment analysis. Subsequently, we could use polarity as a covariate to also capture how the sentiment of the mRNA-related events evolve over time. Although it is not the focus of this study, a closer examination of the users *might* also provide some meaningful information on whether certain users are more likely to post or comment on certain topics (cf. [12]).

Acknowledgements

This work is part of the project Rumour Mining (MXM19-1161:1) financed by the Bank of Sweden Tercentenary Foundation (Riksbankens Jubileumsfond).

References

- [1] Leong C, Jin L, Kim D, et al. Assessing the impact of novelty and conformity on hesitancy towards COVID-19 vaccines using mRNA technology. *Com Med*. 2022, 2:61. doi:10.1038/s43856-022-00123-6
- [2] Aldén M, Olofsson Falla F, et al. Intracellular Reverse Transcription of Pfizer BioNTech COVID-19 mRNA Vaccine In Vitro in Human Liver Cell Line. *Mol. Bio*. 2022, 44, 1115–1126.
- [3] Scannell D, Desens L, Guadagno M, et al. COVID-19 Vaccine Discourse on Twitter. *J Health Com*. 2021, 26:7, 443–459. doi: 10.1080/10810730.2021.1955050
- [4] World Health Organization. Ten threats to global health in 2019, 2019. [cited 2023 Jan 03]. Available from: <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019>
- [5] Rutjens BT, van der Linden S, et al. Science skepticism in times of COVID-19. *Group Processes & Intergroup Relations*. 2021. 24(2), 276–283. doi:10.1177/1368430220981415
- [6] Garrett R, Young SD. Online misinformation and vaccine hesitancy. *Translational Behavioral Medicine*. 2021. 11(12):2194–2199. doi:10.1093/tbm/ibab128
- [7] Klimiuk K, et al. Vaccine misinf. on social media—topic-based content & sentiment analysis of Polish vaccine-deniers' comments on FB. *Hum Vacc&Imm*. 2021. 17:7. doi:10.1080/21645515.2020.1850072
- [8] Ahmed W, Vidal-Alaball J, et al. COVID-19 and the 5G conspiracy theory: Social network analysis of twitter data. *J Med. Int Res*. 2020. 22, e19458. doi:10.2196/19458
- [9] Roberts M, Brandon S, Tingley D. Stm: An R package for structural topic models. *J of Statistical Software*. 2019 91.1: 1–40. doi:10.18637/jss.v091.i02
- [10] Blei DM, et al. Latent Dirichlet Allocation. *J Mach Learn Res*. 2003 3:993–1022. doi:jmlr.2003.3.4-5.993
- [11] Mimno D, Wallach HM, et al. Optimizing Semantic Coherence in Topic Models. *Empirical Methods in NLP*; 2011 July 27–31; Edinburgh, UK, c Association for Comp. Ling., p. 262–272.
- [12] Lyu JC, Le Han E, Luli GK. COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. *J Med Inter Res*. 2021 Jun 29;23(6):e24435. doi:10.2196/24435.