Measuring COVID-19 Vaccine Hesitancy: Consistency of Social Media with Surveys

Ninghan Chen[1], Xihui Chen[1] ✉, Jun Pang[1], Liyousew G. Borga[2], Conchita D'Ambrosio[2], and Claus Vögele[2]

[1] Department of Computer Science, University of Luxembourg, Esch-sur-Alzette, Luxembourg
{ninghan.chen, xihui.chen, jun.pang}@uni.lu
[2] Department of Behavioural and Cognitive Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg
{liyousew.borga, conchita.dambrosio, claus.voegele}@uni.lu

Abstract. We validate whether social media data can be used to complement social surveys to monitor the public's COVID-19 vaccine hesitancy. Taking advantage of recent artificial intelligence advances, we propose a framework to estimate individuals' vaccine hesitancy from their social media posts. With 745,661 vaccine-related tweets originating from three Western European countries, we compare vaccine hesitancy levels measured with our framework against that collected from multiple consecutive waves of surveys. We successfully validate that Twitter, one popular social media platform, can be used as a data source to calculate consistent public acceptance of COVID-19 vaccines with surveys at both country and region levels. In addition, this consistency persists over time although it varies among socio-demographic sub-populations. Our findings establish the power of social media in complementing social surveys to capture the continuously changing vaccine hesitancy in a global health crisis similar to the COVID-19 pandemic.

Keywords: Twitter · surveys · COVID-19 vaccine hesitancy

1 Introduction

The last two and half years have seen the impacts of the unprecedented global COVID-19 pandemic on public health and the economy. Thanks to the successful vaccination program, our societies are gradually reopening and going back to the pre-pandemic states. So far, 68.1% of worldwide populations have been fully vaccinated. This milestone cannot be achieved without fast and accurate understanding of the opinions and responses of general populations towards COVID-19 vaccines and their changes over time. For instance, it allows for identifying the right intervention time and evaluating the effectiveness of deployed measures.

Social media has shown its strengths in complementing conventional surveys to study vaccine hesitancy [27]. Social media overcomes the decreasing response rates of surveys and provides a cost-effective way to reach a significantly larger population [14]. In addition, it allows for capturing the evolution of public opinions over time, especially, in case of emergent incidents such as a sudden outburst of misinformation when there is no sufficient time for conducting surveys. In spite of these advantages, the results derived from social media are often questioned mainly because of three inherent sources of errors: *measurements*, *coding* and *missingness* [22, 4]. Measurement errors are incurred by the fact that social media users may not express their real attitudes in their posts while coding errors come from the deficiency of methods in capturing public opinions. Missingness is caused by non-representative social media users, namely, not all people express their opinions online. For instance, Twitter is more favourable to young users while Facebook attracts the elders [33].

We aim to address these challenges confronted in measuring the levels of public vaccine hesitancy with Twitter, one of the most widely used sources of social media data [34, 27, 10]. Unlike existing works examining correlated factors [27], our purpose is to exemplify that with properly designed methods, individuals' vaccine hesitancy can be *accurately* measured from social media and the estimation is consistent with surveys *continuously over time* and *across countries and regions*. To the best of our knowledge, this is the first attempt to study the temporal consistency of social media with surveys regarding vaccine hesitancy.

We perform a cross-validation by making use of the social survey of multiple waves we conducted and the collected 745,661 tweets related to COVID-19 vaccines from three Western European countries. We take advantage of recent advances in natural language processing techniques, and quantify individuals' vaccine hesitancy based on their attitudes expressed in textual posts. In order to overcome the missingness errors caused by non-representative Twitter users, we show that with three socio-demographic attributes, i.e., gender, age and political ideology, the demographic selection bias can be effectively corrected. When designing our framework, we consider its applicability in a global pandemic like COVID-19 and ensure it can be used in multilingual environments.

With comprehensive analysis, we successfully validate that Twitter is able to give close estimation of vaccine hesitancy to surveys. This closeness persists at a similar level across geographical regions and over time. The large Pearson correlation coefficients indicate at least a strong correlation between the results from surveys and Twitter. We also show that the consistency varies among different socio-demographic groups. Our research re-established the power of Twitter to act as a complementary source to continuously monitor public vaccine hesitancy in COVID-19 and future health crises of similar types.

2 Related Work

Since the outbreak of the COVID-19 pandemic, great efforts have been devoted to studying the potential of social media in understanding the public's hesitancy in the fast developed vaccines [10, 39, 2, 36], based on the pre-pandemic success in studying public opinions [34, 23, 27, 30]. For instance, Cascini et al. [10] reviewed the literature during the COVID-19 pandemic about how diffused information on social media impacts vaccination attitudes. In general, previous works aim to study the correlation of social media users' online activities, e.g., information perception, to vaccine hesitancy. According to sources of online digital traces, the related work falls into two categories. The first category makes use of questionnaires or public polls to collect participants' usage habits of various social media platforms as well as their vaccination attitudes. For instance, with a survey of 504 participants, Alfatease et al. [2] observed the dependence between social media usage and willingness to accept vaccination in Saudi Arabia. Wilson et al. [39] revealed the correlation between online disinformation campaign and activity organisation on social media to vaccine hesitancy. The second category leverages tools such as stance detection to infer various features of online activities from social media data of various formats including hashtags, hyperlinks and textual posts. For instance, Shaaban et al. [36] studied vaccine acceptance with positions and tones of comments on various social media platforms. Lyu et al. [25] inferred user demographics as well as vaccine attitudes through a text-based machine learning approach, and analysed vaccine acceptance among people with different demographic characteristics.

Three characteristics have been well accepted as the advantages of social media over surveys, i.e., *volume*, *velocity* and *variety* [32] and promoted social media data as a complementary or alternative source of public opinions. However, the inherent limits such as the bias of population coverage and accuracy of extracted opinions, inevitably cause doubts about claims drawn from social media [35]. Several attempts have been conducted to study the reliability of social media data in studying public opinions by comparison to surveys [15, 34, 27, 3]. Davis et al. [15] compared the sentiments of tweets to the polls about public opinions of the Obamacare act and showed the comparability of Twitter public opinions with survey results. Scarborough [34] illustrated the correlation of tweet sentiments to gender attitudes. Amaya et al. [3] evaluated three types of errors that generate the difference between social media and public polls.

Identified challenges. Few existing works study how and whether individuals' vaccine hesitancy can be directly estimated with digital traces on social media, and whether the estimation is consistent with surveys, especially over time. Although a number of factors have been revealed to be correlated, they can only be interpreted as indicators but not a precise estimation. Without a proper cross-validation, it is unclear whether social media can be used for real-time vaccine hesitancy monitoring as suggested [31]. The work most related to this paper is [18], which compares existing selection bias correction methods with demographic attributes extracted with machining learning models. Different from our paper, it aims at public health status and does not study the consistency of the predictability of online discourses over time.

3 Survey and Twitter Data

Survey. We conducted a survey of people over 18 years of age in 6 European countries, ensuring in each country that the sample was representative in terms of

Table 1. Statistics of survey participants and Twitter users.

	#Survey participants	**#Twitter users**	**#Tweets**
Luxembourg	474	1,764	28,148
Germany	501	13,390	270,695
France	711	26,562	446,818

gender, region and age. Information on the status of respondents in the pandemic was collected in order to study the impact of the COVID-19 pandemic [8].

We select three adjacent countries, i.e., Germany, France and Luxembourg, as our research objects because of their synchronised vaccination policies and close economic connections. Moreover, the diverse origins of the people are also representative for the worldwide populations. Respondents were invited to fill in online questionnaires including questions about their living conditions, mental health and opinions about vaccination. Meanwhile, socio-demographic characteristics such as age, gender, education and income are also collected. Our survey is conducted in multiple waves at intervals of approximately 4 months. During the waves in June and October of 2021, and March 2022, we consecutively asked about individuals' vaccination attitudes through the following question:

Have you been vaccinated against COVID-19?
1 Yes
2 No, but I plan to
3 No and I do not plan to

More than 8,000 individuals participated in the first wave. However, only part of them participated consistently in the following waves. As one of our purposes is to test whether Twitter can capture the changes in individuals' vaccination attitudes over time, we only keep the participants that responded in all the three waves. Table. 1 shows the statistics of our survey data.

Vaccine hesitancy evaluation. The vaccine hesitancy is calculated as the proportion of the participants marking the third option. Let $N_i^{\ell,t}$ $(i \in \{1, 2, 3\})$ be the number of respondents from a given region/country ℓ ticking the i-th option, in a given survey wave t. As the first two options indicate acceptance of COVID-19 vaccines, the vaccine hesitancy of a region ℓ in the survey wave t, denoted by VH_ℓ^t, is calculated as: $VH_\ell^t = \frac{N_3^{\ell,t}}{N_1^{\ell,t}+N_2^{\ell,t}+N_3^{\ell,t}}$.

Twitter Data. We constructed a dataset of Twitter users located in our targeted countries who actively participated in vaccine-related discussions in the periods corresponding to the selected three survey waves. Their tweets are also needed to infer their vaccine hesitancy. As we will see later in Section 4.1, in order to employ the vaccination attitude learning model [13], we crawled their social connections as well. Instead of directly crawling tweets worldwide, for the purpose of efficiency, we referred to a publicly available Twitter dataset [11] to obtain the preliminary set of users. The dataset consists of the IDs of 2,198,090 tweets related to vaccination originating from four European countries, i.e., Germany, France, Belgium and Luxembourg generated up to March 2021. Among these tweets, about 17,934 are annotated with the vaccination attitudes expressed, i.e., *positive*, *negative* and *neutral*. We obtained the tweets according to the published tweet IDs with the official Twitter API. From the associated metadata of every downloaded tweet, we derive the IDs of the originating user and his/her location. The geographic information of a tweet is either self-provided by the originator or attached by the device's positioning services such as GPS. We adopted ArcGis, the same approach used in previous works [20], to regularise the locations whenever they are ambiguous into the form of countries and regions. When a user posted multiple tweets with different regions, we select the most frequently used one as the user's location. We only kept the 49,791 users located in Luxembourg, Germany and France. We further downloaded the following relations of each user and constructed a social network represented as a directed graph. A vertex represents a Twitter user while an edge from vertex v to vertex v' indicates that the user corresponding to v follows the user represented by v'.

With the identified Twitter users, we downloaded their tweets posted in the three months when the targeted survey waves were conducted. We used the same keywords as [11] to filter the tweets related to COVID-19 vaccines. We only kept users that posted at least 5 tweets in every targeted month to ensure the reliability of vaccination attitudes calculation. Note that we do *not* consider retweets because compared to quoted and original tweets, they are more likely to carry the intentions of their originating users. As Twitter contains accounts maintained by organisations such as newsagents and healthcare departments, we removed such organisation accounts to ensure that vaccination attitudes belong to the general population. We applied the pre-trained model in [38] to identify such accounts. In total, we removed $5,070$ organisation accounts. In the end, we have $1,764$ Twitter users from Luxembourg, $13,390$ from Germany and $26,562$ from France, which are almost 30 times as many as the survey respondents. The IDs of our collected tweets are available at `https://anonymous.4open.science/r/country_3_vax_data-43F5/`.

4 Measuring Vaccine Hesitancy with Twitter

We select Twitter as the source of vaccination attitudes by assuming Twitter users tend to express their real opinions about COVID-19 vaccines. In other words, we hypothesise the *measurement* error is acceptable when Twitter data is used. In this section, we describe how we handle the other two inherent errors with three sequential steps. The first step targets at reducing the coding error by proposing a measurement of vaccine hesitancy while the other two steps are to correct the missingness error. Note that our aim is not to eliminate the errors but to mitigate the impact of these errors. We adopt widely accepted methodologies to avoid statistic manipulation and thus ensure the generality of our framework.

4.1 Measuring individual vaccine hesitancy

A significant amount of research has been devoted in understanding public opinions from social media posts, varying from word-level [19, 6] to data-driven approaches [28, 41]. We use one recent deep learning model which is specifically designed to infer COVID-19 vaccination attitudes expressed in tweets and overwhelms existing models in classification accuracy [13]. Another reason for our selection is its power of dealing with multilingualism which is essential for the global demands of vaccination attitude monitoring. Intuitively, the model uses RoBERTa [29], the most popular pre-trained embedding method, to calculate the representation of tweets, and leverages social connections to integrate the recent tweets of each user's friends with a variant of H2GCN [42]. The model takes the text representation of a tweet under analysis and the integrated embedding of the recent discourse of the originating user's friends as input and output the possibility that the tweet is classified into attitudes corresponding to *vaccine support*, *anti-vaccine*, and *neutral*. We retrain this model with the release annotations [11], our constructed social network and collected tweets. The resulted model achieves an accuracy of 0.80 and Marco F1-score of 0.79.

Vaccine hesitancy calculation. To estimate an individual user's vaccine hesitancy, we leverage the vaccination attitudes expressed in his/her tweets. Inspired by the measurements in [7, 12] which are originally proposed to evaluate subjective well-being, we construct the measurement of vaccine hesitancy. Intuitively, users who post more tweets supporting vaccination are considered more acceptable of COVID-19 vaccines and thus more likely to get vaccinated. Formally, let $N_s(u)$, $N_a(u)$ and $N_{neu}(u)$ be the numbers of tweets of user u posted in a given period t (i.e., June and October 2020, and March 2022 in our analysis), indicating his supportive, anti-vaccine and neutral stance about COVID-19 vaccination, respectively. The vaccine hesitancy of u, denoted by $VH^t(u)$, is calculated as:

$$VH^t(u) = 1 - \frac{VA^t(u)+1}{2}, \text{ where } VA^t(u) = \frac{N_s(u)-N_a(u)}{N_s(u)+N_a(u)} \cdot \left(\frac{N_s(u)+N_a(u)}{N_s(u)+N_a(u)+N_{neu}(u)}\right)^{\frac{1}{2}}. \quad (1)$$

Note that $VA^t(u)$ is extended by Chen et al. [12] from [7] with *neutral* messages considered by adding a scaling factor. It actually measures the vaccination acceptance of user u and is between -1 and 1. We first normalise it to the domain between 0 to 1 and then compute the complement as the level of vaccine hesitancy. As a result, a user's vaccine hesitancy of 1 indicates total opposition against vaccination and 0 means complete belief in COVID-19 vaccines.

4.2 Inferring socio-demographic profiles

The missingness error is related to the socio-demographic selection bias which is a well-recognised inherent limit of social media [33]. One way to correct such bias is to adjust each individual's vaccine hesitancy level by multiplying a factor that is calculated according to the difference between the distributions of social media users and the general population. Despite the large number of out-of-box methods inferring various demographic attributes such as education [18] and income [26], few can be used in our analysis due to their limitation in dealing with multilingual texts. Existing methods, especially the ones based on machine/deep learning, can be extended to multilingual data with well-annotated multilingual data for training and testing. However, due to the privacy protection regulations such as GDPR, it is challenging to collect people's social media accounts and their corresponding socio-demographic information.

In order to ensure our framework applicable globally, we need to select the demographic attributes that can be inferred with multilingual data and effectively mitigate socio-demographic selection bias. Considering these two requirements, we select three socio-demographic attributes, i.e., age, gender and political ideology. We detail the models adopted or extended to infer these three attributes.

Age and gender. We use the multi-modal deep neural model M3 [38] to infer the age and gender of Twitter users. These two attributes are simultaneously inferred by M3 with users' account name, screen name, self-descriptive description and profile image. A user's age falls into one of the three ranges: $19-29$, $30-39$ and ≥ 40. Multilingual textual inputs are first translated into English word by word, and the $3,000$ most frequent characters are selected to calculate users' embedding. Although the performance of the M3 model has been confirmed by previous studies [41, 17], we still construct a sample dataset to test its performance on our collected Twitter data. This sample dataset consists of 100 randomly selected users, whose ages and genders are manually annotated by two annotators. The annotated labels are highly agreed between the two annotators with large Cohen's Kappa coefficients ($k = 0.95$ for gender and $k = 0.81$ for age). When tested on our sample dataset, the M3 model achieves a Macro F1 score of 0.92 and an accuracy of 0.91 for age classification. For gender classification the Macro F1 score is 0.78 and the accuracy is 0.75.

Political ideology. We infer Twitter users' political ideology by the Multi-task Multi-relational Embedding model (TIMME) [40]. TIMME utilises the heterogeneous types of relationships between Twitter users including '*following*', '*retweet*', '*mention*' and '*like*' in conjunction with tweets to infer users' bipolar political ideologies, i.e., left and right. As TIMME is designed for English-only data, we have to re-train it on a multilingual dataset. One distinguishing feature of TIMME is that it can be trained only with a sparsely annotated training set. This allows us to prepare a new training set of a relatively small size from our collected Twitter data with the public Twitter parliamentarians dataset [37]. The dataset [37] contains manually verified parliamentarians from 26 countries, including France, Germany, and Luxembourg, with their names, political ideology, and Twitter IDs. The political ideology is evaluated in a scale from 0 to 10. We manually update the list of parliamentarians of the three countries by i) adding new politicians who joined after the data release, and ii) updating the obsolete Twitter IDs. In total, we constructed a training dataset of 1,021 parliamentarians. We encode the political ideology scores into left, centre and right. Specifically, a score smaller than 4 is encoded as left while a score larger than 6 is encoded as right. A score between 4 and 6 belongs to centre.

We conduct two extensions to TIMME. First, we extend TIMME to a triple classification model (with '*centre*' added) by replacing the binary cross-entropy loss function with a categorical cross-entropy loss function. Second, to enable TIMME to handle multilingual texts, we replace the word-level embedding with

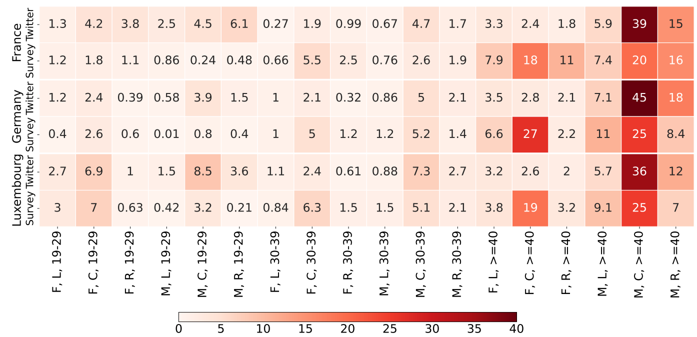


Fig. 1. Population distribution according to age, gender (F: female, M: male) and political ideology (L: left, C: centre, R: right).

RoBERTa [29]. We train the extended TIMME model with the parliamentarian dataset and achieve an accuracy of 0.77 and Marco F1-score of 0.78.

Socio-demographic selection bias in our Twitter data. Figure 1 presents the socio-demographic distributions of the survey participants and our collected Twitter users in the three targeted countries. A significant difference between the two distributions in every country is observed. Moreover, the difference varies from one country to another. When measured by KL-Divergence, we have the distances of 0.52, 0.29 and 0.38 in France, Germany and Luxembourg. This shows the non-representation of Twitter users and the necessity of correction.

4.3 Correcting socio-demographic selection bias

The general idea of socio-demographic correction is to re-weigh non-representative samples' vaccine hesitancy with scalars calculated according to their percentage differences from the representative population. Let ϕ_u be the socio-demographic attributes of user u in the form of age, gender and political ideology. Suppose $\Pr_\ell^S(\phi_u)$ ($\Pr_\ell^T(\phi_u)$) be the percentage of survey participants (Twitter users) with the same demographic attribute as u in region ℓ. We use $\mathcal{U}^\ell$ to denote the set of users located in the given region ℓ. Thus, the corrected average vaccine hesitancy of region ℓ in time period t is

$$\widehat{VH}_\ell^t = \frac{1}{|\mathcal{U}^\ell|} \sum_{u \in \mathcal{U}^\ell} VH^t(u) \cdot \frac{\Pr_\ell^S(\phi_u)}{\Pr_\ell^T(\phi_u)}. \tag{2}$$

According to the availability of the joint distributions (i.e., $\Pr_\ell^S$ and $\Pr_\ell^T$), we can use different implementations. When the joint distributions are available, the correction is called *post-stratification*. When the two joint distributions are not both available, *naive post-stratification* [24] and *Raking* [16] are applicable. The former assumes independent socio-demographic variables while Raking adopts an iterative approach to adjust each sample's marginal to match the representative population distribution.

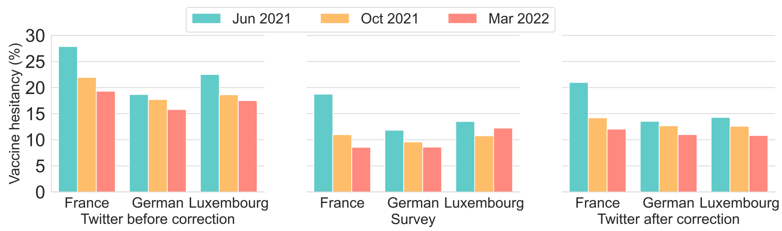


Fig. 2. Vaccine hesitancy across countries.

5 Cross-validation

Our objective of the cross-validation is to test whether the vaccine hesitancy inferred from Twitter with our framework is similar to that collected from the survey and whether the similarity, if validated, persists over time and across regions/countries. As vaccine hesitancy varies among countries and regions [21], we study both the country- and region-level vaccine hesitancy. Note that we use post-stratification as the correction method because of the availability of the joint distributions of the three selected socio-demographic attributes.

5.1 Vaccine hesitancy across countries

In Figure 2, we show the average vaccine hesitancy in Germany, France and Luxembourg in the three survey waves calculated with Twitter data and surveys. In general, Twitter users are more negative about vaccination. In addition, we have three other observations. First, we observe similar changes of vaccine hesitancy over time. This complies with the latest updates derived from surveys/polls around the world which indicate a decreasing trend of vaccine hesitancy [9, 5]. This trend presents in all the three countries even without the socio-demographic selection bias correction. Special attention should be paid to the survey of Luxembourg in the last wave. The vaccine hesitancy increased by about 0.015 compared to the second wave. With a manual check, we notice that about 8 participants changed their choice from '*No but I plan to*' to the third option '*No and I do not plan to*'. This increase is actually not consistent with the continuous increase of vaccinated population since October 2021 and may be caused by the relatively smaller number of respondents in Luxembourg. Second, when ordered by their vaccine hesitancy, the countries have similar rankings. Residents in France are relatively more reluctant to get vaccinated compared to the other two countries and people in Germany are more favourable to vaccination. Our third observation is that without bias correction, the vaccine hesitancy calculated with Twitter data is rather different from the survey while correcting selection bias can significantly reduce the difference and ensure a similar estimation. Without the bias correction, the average vaccine hesitancy differences of the three countries are 0.083, 0.089 and 0.077 in the three waves, respectively. The differences drop by more than 70% to 0.019, 0.027 and 0.024 after correction.

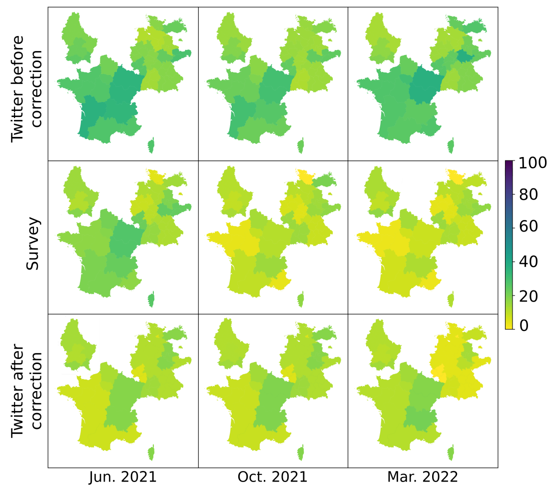
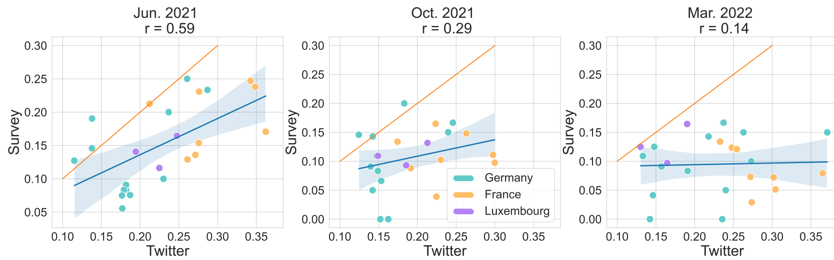


Fig. 3. Vaccine hesitancy across regions from Twitter and survey.

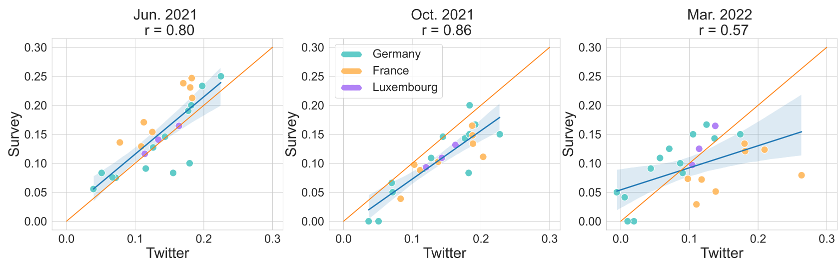
5.2 Vaccine hesitancy across regions

We obtain the regions according to the administrative divisions of the three countries. As the distribution of our survey respondents is not uniform across regions, to ensure the reliability of the vaccination reluctance calculated from our survey, we remove those regions with fewer than 11 participants. In total, we obtain 24 regions including 8 administrative regions of France and 13 states of Germany. Due to the small size of Luxembourg communities, we divide Luxembourg into three regions: north, south and central.

Figure 3 illustrates the region-level vaccine hesitancy in the three survey waves. We can clearly see that after bias correction, Twitter data can reflect similar levels of vaccine hesitancy to the surveys despite the relatively big differences in certain regions. This similarity persists in all the three waves. In Figure 4, we further show the Pearson correlation coefficient r between the hesitancy levels calculated from Twitter and surveys before and after the socio-demographic bias correction. Each point corresponds to a region with a coordinate (x, y) where x is the vaccine hesitancy derived from Twitter and y is that from the surveys. The orange line is composed by the points where $x = y$. After correction, the Pearson correlation coefficients reach over 0.80 in the first two waves, which indicates a *very strong* correlation according to the well-accepted standard [1]. In the third wave, the correlation strength decreases to 0.57 which is still interpreted as *strong*. After a closer look, we observe that the points that are relatively far from the orange line mainly belong to France and Twitter users acted more negatively about COVID-19 vaccines. We check the news in March 2022 and find that this is the period when the Omicron variants were transmitted fast in eastern France regions. This also implies that Twitter can capture the changes of vaccination attitudes faster than survey in emergent incidents.



(a) Twitter vs. Survey before bias correction ($p < 0.05$).



(b) Twitter vs. Survey after bias correction ($p < 0.05$).

Fig. 4. Region-level correlations of vaccine hesitancy between Twitter and Survey.

We further check whether the above identified correlation persists in various socio-demographic sub-populations. We divide the survey respondents and Twitter users into 18 groups according to their age, gender and political ideology. Then we calculate the region-level vaccination reluctance rates for each group, and compute the Pearson correlation between the reluctance rates of each Twitter group and the corresponding survey group. Figure 5 depicts the results. The general observation is that the correlation indeed varies among different demographic groups. The correlation increases for groups with larger ages but remains almost the same regardless of gender and political ideology. This implies that younger people may actively participate in discussion about vaccines, but they are less willing to express their real intention of vaccination on Twitter. In addition, the correlation decreases with time, which implies when a high-level vaccination rate is reached, the topics on Twitter about COVID-19 vaccines become less relevant to users' intention of vaccination.

6 Discussion and Conclusion

We have proposed a framework to directly estimate public vaccine hesitancy from Twitter. Our framework addressed the widely recognised inherent errors when analysing social media data with a quantitative measurement of vaccine hesitancy and an adapted method correcting socio-demographic selection bias. With our multi-wave surveys and collection of tweets, we conducted the first at-

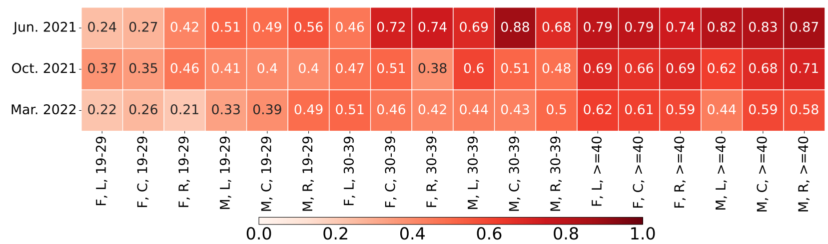


Fig. 5. Pearson correlations (r) between Twitter and survey across 24 regions by Age, gender (F: female, M: male) and political ideology (L: left, C: centre, R: right).

tempt to validate the consistency between Twitter and surveys regarding monitoring COVID-19 vaccine hesitancy both across regions and over time. Through comprehensive cross-validation, we have shown that Twitter can capture the public vaccine hesitancy and generate at least strongly correlated estimation with that inferred by surveys. Moreover, this correlation is consistent over time on levels of both countries and regions although it varies among different socio-demographic sub-populations. Last but not least, we considered the global demands of vaccination attitude monitoring and empower our framework to deal with multilingual texts. With this paper, we re-established the power of social media in complementing social surveys to continuously capture the fast evolution of vaccine hesitancy in public health crises like COVID-19. Moreover, our work can encourage social scientists to use social media in studies, especially for the topics which are hard to formulate in questionnaires e.g., influences of social interactions on vaccine hesitancy.

We have a few limitations that will be addressed in future. First, our cross-validation is conducted in Western Europe. Similar studies in other areas can further validate the generality of our framework and our findings. Second, with vaccine hesitancy consistency validated, it will be helpful to examine whether existing social findings such as correlated factors can also be confirmed on social media data. Third, we only tested three socio-demographic attributes. In spite of their effectiveness in bias correction, other socio-demographic attributes can also be tested and added to the bias correction if they can lead to better performance, especially with new progress in artificial intelligence.

Ethical consideration. This work is solely grounded in public data and does not involve any private information from individuals. The research process was established in full compliance with FAIR data principles, Twitter Developer Agreement & Policy and relevant policies. The survey is also approved by the Ethics Review Panel of our institution.

Acknowledgement. This research was funded in whole, or in part, by the Luxembourg National Research Fund (FNR), grant reference PRIDE17/12252781 (DRVIVEN), C21/IS/16281848 (HETERS) and 14840950 (COME-HERE). The research is also supported by André Losch Fondation, Art2Cure, Cargolux, and CINVEN Foundation.

References

1. Akoglu, H.: User's guide to correlation coefficients. Turkish Journal of Emergency Medicine **18**(3), 91–93 (2018)
2. Alfatease, A., Alqahtani, A.M., Orayj, K., Alshahrani, S.M.: The impact of social media on the acceptance of the COVID-19 vaccine: a cross-sectional study from Saudi Arabia. Patient Prefer Adherence **15**, 2673–2681 (2021)
3. Amaya, A., Bach, R., Kreuter, F., Keusch, F.: Measuring the strength of attitudes in social media data, chap. 5, pp. 163–192. John Wiley & Sons, Ltd (2020)
4. Baker, R.: Big Data, chap. 3, pp. 47–69. John Wiley & Sons, Ltd (2017)
5. Bartoš, V., Bauer, M., Cahlíková, J., Chytilová, J.: Communicating doctors' consensus persistently increases COVID-19 vaccinations. Nature **606**(7914), 542–549 (2022)
6. Bi, D., Kong, J., Zhang, X., Yang, J.: Analysis on health information acquisition of social network users by opinion mining: case analysis based on the discussion on COVID-19 vaccinations. Journal of Healthcare Engineering **2021** (2021)
7. Bollen, J., Gonçalves, B., Ruan, G., Mao, H.: Happiness is assortative in online social networks. Artificial Life **17**(3), 237–251 (2011)
8. Borga, L.G., Clark, A.E., D'Ambrosio, C., Lepinteur, A.: Characteristics associated with COVID-19 vaccine hesitancy. Scientific Reports **12**(1), 1–9 (2022)
9. Burger, R., Köhler, T., Golos, A.M., Buttenheim, A.M., English, R., Tameris, M., Maughan-Brown, B.: Longitudinal changes in COVID-19 vaccination intent among South African adults: evidence from the NIDS-CRAM panel survey, February to May 2021. BMC Public Health **22**(1), 1–10 (2022)
10. Cascini, F., Pantovic, A., Al-Ajlouni, Y.A., Failla, G., Puleo, V., Melnyk, A., Lontano, A., Ricciardi, W.: Social media and attitudes towards COVID-19 vaccination: a systematic review of the literature. eClinicalMedicine **48**, 101454 (2022)
11. Chen, N., Chen, X., Pang, J.: A multilingual dataset of COVID-19 vaccination attitudes on Twitter. Data in Brief **44**, article 108503 (2022)
12. Chen, N., Chen, X., Zhong, Z., Pang, J.: The burden of being a bridge: analysing subjective well-being of Twitter users during the COVID-19 pandemic. In: Proc. 2022 European Conference on Machine Learning and 26th Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD). Lecture Notes in Computer Science, Springer (2022)
13. Chen, N., Chen, X., Zhong, Z., Pang, J.: "Double vaccinated, 5G boosted!": learning attitudes towards COVID-19 vaccination from social media (2022). https://doi.org/10.48550/ARXIV.2206.13456
14. Chen, N., Chen, X., Zhong, Z., Pang, J.: Exploring spillover effects for COVID-19 cascade prediction. Entropy **24**(2), 222 (2022)
15. Davis, M.A., Zheng, K., Liu, Y., Levy, H.: Public response to obamacare on Twitter. Journal of Medical Internet Research **19**(5), e167 (2017)
16. Deville, J.C., Särndal, C.E., Sautory, O.: Generalized raking procedures in survey sampling. Journal of the American statistical Association **88**(423), 1013–1020 (1993)
17. Duong, V., Luo, J., Pham, P., Yang, T., Wang, Y.: The ivory tower lost: how college students respond differently than the general public to the COVID-19 pandemic. In: Proc. 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). pp. 126–130. IEEE (2020)
18. Giorgi, S., Lynn, V.E., Gupta, K., Ahmed, F., Matz, S., Ungar, L.H., Schwartz, H.A.: Correcting sociodemographic selection biases for population prediction from

social media. Proc. 2022 International AAAI Conference on Web and Social Media (ICWSM) **16**(1), 228–240 (2022)
19. Gori, D., Durazzi, F., Montalti, M., Di Valerio, Z., Reno, C., Fantini, M.P., Remondini, D.: Mis-tweeting communication: a vaccine hesitancy analysis among Twitter users in Italy. Acta Bio Medica: Atenei Parmensis **92**(S6) (2021)
20. Hecht, B.J., Hong, L., Suh, B., Chi, E.H.: Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. In: Proc. 2011 International Conference on Human Factors in Computing Systems (CHI). pp. 237–246. ACM (2011)
21. Hou, Z., Tong, Y., Du, F., Lu, L., Zhao, S., Yu, K., Piatek, S.J., Larson, H.J., Lin, L., et al.: Assessing COVID-19 vaccine hesitancy, confidence, and public engagement: a global social listening study. Journal of medical Internet research **23**(6), e27632 (2021)
22. Hsieh, Y.P., Murphy, J.: Total Twitter error, chap. 2, pp. 23–46. John Wiley & Sons, Ltd (2017)
23. Jaidka, K., Giorgi, S., Schwartz, H.A., Kern, M.L., Ungar, L.H., Eichstaedt, J.C.: Estimating geographic subjective well-being from Twitter: a comparison of dictionary and data-driven language methods. Proceedings of the National Academy of Sciences **117**(19), 10165–10171 (2020)
24. Leemann, L., Wasserfallen, F.: Extending the use and prediction precision of subnational public opinion estimation. American Journal of Political Science **61**(4), 1003–1022 (2017)
25. Lyu, H., Wang, J., Wu, W., Duong, V., Zhang, X., Dye, T.D., Luo, J.: Social media study of public opinions on potential COVID-19 vaccines: informing dissent, disparities, and dissemination. Intelligent Medicine **2**(1), 1–12 (2022)
26. Matz, S.C., Menges, J.I., Stillwell, D.J., Schwartz, H.A.: Predicting individual-level income from Facebook profiles. PLOS ONE **14**(3), e0214369 (2019)
27. Nowak, S.A., Chen, C., Parker, A.M., Gidengil, C.A., Matthews, L.J.: Comparing covariation among vaccine hesitancy and broader beliefs within Twitter and survey data. PLOS ONE **15**(10), 1–16 (10 2020)
28. Nyawa, S., Tchuente, D., Fosso-Wamba, S.: COVID-19 vaccine hesitancy: a social media analysis using deep learning. Annals of Operations Research **16**, 1–39 (2022)
29. Ou, X., Li, H.: Ynu_oxz @ haspeede 2 and AMI : XLM-RoBERTa with ordered neurons LSTM for classification task at EVALITA 2020. In: Proc. 2020 Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA). vol. 2765. CEUR-WS.org (2020)
30. Piedrahita-Valdés, H., Piedrahita-Castillo, D., Bermejo-Higuera, J., Guillem-Saiz, P., Bermejo-Higuera, J.R., Guillem-Saiz, J., Sicilia-Montalvo, J.A., Machío-Regidor, F.: Vaccine hesitancy on social media: sentiment analysis from June 2011 to April 2019. Vaccines **9**(1) (2021)
31. Puri, N., Coomes, E.A., Haghbayan, H., Gunaratne, K.: Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Huamn Vaccines & Immunotherapeutics **16**(11), 2586–2593 (2020)
32. Reveilhac, M., Steinmetz, S., Morselli, D.: A systematic literature review of how and whether social media data can complement traditional survey data to study public opinion. Multimedia Tools and Applications **81**, 10107–10142 (2022)
33. Ribeiro, F.N., Benevenuto, F., Zagheni, E.: How biased is the population of Facebook users? Comparing the demographics of Facebook users with census data to generate correction factors. In: Proc. 2020 ACM Conference on Web Science (WebSci). pp. 325–334. ACM (2020)

34. Scarborough, W.J.: Feminist Twitter and gender attitudes: opportunities and limitations to using Twitter in the study of public opinion. Socius **4**, 2378023118780760 (2018)
35. Schober, M.F., Pasek, J., Guggenheim, L., Lampe, C., Conrad, F.G.: Social media analyses for social measurement. Public Opinion Quarterly **80**(1), 180–211 (2016)
36. Shaaban, R., Ghazy, R.M., Elsherif, F., et al.: COVID-19 vaccine acceptance among social media users: a content analysis, multi-continent study. International Journal of Environmental Research and Public Health **19**(9) (2022)
37. van Vliet, L., Törnberg, P., Uitermark, J.: The Twitter parliamentarian database: analyzing Twitter politics across 26 countries. PLOS ONE **15**(9), e0237073 (2020)
38. Wang, Z., Hale, S.A., afnd Przemyslaw A. Grabowicz, D.I.A., Hartmann, T., Flöck, F., Jurgens, D.: Demographic inference and representative population estimates from multilingual social media data. In: Proc. 2019 The World Wide Web Conference (WWW). pp. 2056–2067. ACM (2019)
39. Wilson, S.L., Wiysonge, C.: Social media and vaccine hesitancy. BMJ Global Health **5**(10) (2020)
40. Xiao, Z., Song, W., Xu, H., Ren, Z., Sun, Y.: Timme: Twitter ideology-detection via multi-task multi-relational embedding. In: Proc. 2020 ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD). pp. 2258–2268. ACM (2020)
41. Zhang, C., Xu, S., Li, Z., Liu, G., Dai, D., Dong, C., et al.: The evolution and disparities of online attitudes toward COVID-19 vaccines: year-long longitudinal and cross-sectional study. Journal of Medical Internet Research **24**(1), e32394 (2022)
42. Zhu, J., Yan, Y., Zhao, L., Heimann, M., Akoglu, L., Koutra, D.: Beyond homophily in graph neural networks: Current limitations and effective designs. In: Proc. 2020 Annual Conference on Neural Information Processing Systems (NeurIPS). vol. 33, pp. 7793–7804. NeurIPS (2020)