Comment on “The COVID-19 infodemic does not affect vaccine acceptance”

Riccardo Gallotti,^{1,*} Federico Pilati,^{2,*} Valeria D’Andrea,¹ Francesco Valle,¹ Anna Bertani,^{3,1} Guido Caldarelli,⁴ Pier Luigi Sacco,^{5,†} and Manlio De Domenico^{1,‡}

¹*Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo (TN), Italy*

²*IULM University, Via Carlo Bo 1, 20143 Milan, Italy*

³*Department of Mathematics, University of Trento, 38123 Povo (TN), Italy*

⁴*DSMN, Universit’a Ca’ Foscari, Mestre (VE), Italy*

⁵*IULM University, Via Carlo Bo, 1, 20143 Milan, Italy*

(Dated: July 22, 2021)

In a recent paper by Valensise et al [1], the authors present an analysis of social media data – from Facebook and Twitter – and vaccine hesitancy data – from Facebook – to provide evidence that the overabundance of potentially unreliable information, known as *infodemic*, does not affect vaccine acceptance. If confirmed, this result could have a dramatic impact on public health policies across the world, suggesting that current actions taken in place to contain and prevent the spreading of disinformation and misinformation might be useless to significantly hinder vaccine hesitancy. We disagree with this conclusion on the basis of existing literature that the authors fail to consider, of methodological concerns that suggest that their approach might have crucial flaws, and of an alternative empirical analysis unraveling a broader and richer picture to interpret.

Simplistic analyses are not enough to assess the complex interplay between two complex social and behavioral phenomena such as vaccine hesitancy and infodemic: more sophisticated analyses are needed to account for the different intervening socio-cultural, behavioral, environmental and epidemiological factors. Under these conditions, we conclude that the authors’ main claim is conceptually and empirically unsupported. We are sincerely concerned that, if measures disregarding the circulating disinformation around the COVID19 vaccines were endorsed by policy makers in the design of future public health policies, it might lead to serious negative consequences by dangerously overlooking a major potential driver of dysfunctional behavioral responses to public health policies and goals.

A. CONCEPTUAL CONCERNS

Vaccine hesitancy and social media.— There is ample debate as to whether exposure to the social media has an effect on actual beliefs, attitudes and choices of people in a number of highly relevant social issues. In the current pandemic crisis, vaccination is a particularly hot topic in this regard. Vaccine hesitancy is a long-standing public health issue, whose history dates back to the discovery of the first vaccine [2]. The topic has become, if possible, more relevant than ever during the current crisis. The COVID-19 pandemic has posed an unprecedented challenge to our capacity to develop, in extremely short time, new vaccines that could be reasonably effective against a still poorly known disease agent. Nevertheless, the response has been impressive, with at least six vaccines already approved by many public health authorities around the globe, and some more now entering the approval phase. However, this remarkable accomplishment can be at least partially jeopardized if a significant share of the population refuses to be vaccinated. One can therefore regard vaccine hesitancy as a key variable in determining the overall efficacy of the vaccination strategy, keeping in mind that the relation-

ship between vaccine hesitancy and the actual decision to vaccinate is mediated by a number of concurrent factors of various nature [3]. It would be misleading to think of vaccine hesitancy as a binary phenomenon, with the population divided into favorable and contrarians [4]. It is more appropriate to consider it as a full spectrum of attitudes, ranging from full support to vaccination to irrevocable refusal, but with all possible intermediate nuances. For this reason, only focusing on the most visible and polarized attitudes may be counterproductive for a real understanding of the phenomenon and for an assessment of the most effective policy approaches. Much of the debate on whether or not vaccine hesitancy is sensitive to social media relies on evidence based upon the presumption that exposure to misleading or false content may more or less immediately reflect in corresponding attitudes and choices. However, such presumption is surely debatable.

The most likely effect of social media exposure, and especially so for topics that elicit a strong sense of anxiety while at the same time being characterized by a high variety of strongly conflicting information and viewpoints, is going to be appreciated over relatively long time spans, as repeated exposure to certain contents gradually builds an overarching cognitive frame in which people increasingly identify themselves. Such framing effect is particularly critical for the large share of people who are not pre-judicially hostile to vaccines, but are understandably confused and perplexed by the flood of contradictory, emotionally charged information and claims they are con-

* These two authors contributed equally

† Corresponding author: pierluigi.sacco@fas.harvard.edu

‡ Corresponding author: mdedomenico@fbk.eu

stantly exposed to. Building cognitive frames takes time, and if one evaluates such processes on too short time spans would easily arrive at misleading conclusions. For instance, if one would assess the importance of education by considering an exposure to educational contents of a few years, it could be easy to conclude that education fails to deliver on a variety of crucial goals. However, once the educational process is assessed along an appropriate, longer time span, its effects become apparent. We should not forget that social media have only appeared in the last twenty years, and many of their socio-cognitive effects will only become clear in the years to come. Such considerations should be kept in mind when evaluating the effects of social media exposure on sensitive topics, and special caution should be exerted on a topic such as vaccination in the middle of a pandemic. What is at stake is not only how we respond to the current pandemic crisis, but also what kind of public opinion will be facing the possible future crises and the public opinion on other types of vaccination campaigns such as those for Measles, Chickenpox, etc.

It is understandable that, in view of the urgency of policy responses to the pandemic crisis, there may be an incentive to ask whether new, rapidly growing social phenomena such as infodemics have a real impact on human behaviors on key matters such as vaccine acceptance. However, one must be aware that when studying a still poorly understood social phenomenon on the basis of data from a short time span, no matter how large the sample size, a particularly careful and rigorous methodological approach is needed. In a recent paper, Valensise et al. [1] carry out such an analysis, and claim that social media driven-infodemic has no measurable influence on vaccine acceptance. In this comment, we question their results both on conceptual and methodological grounds, on the basis of the above mentioned considerations. Their results, which runs counter an established theoretical and empirical literature, is largely due to the methodological limitations of their analysis. Alternative analyses, addressing the methodological concerns that we raise, fail instead to replicate their results.

Social media effects on vaccine hesitancy.— To date, studies regarding the impact of an overabundance of health information are few, with even fewer reports regarding the most recent example of the COVID-19 epidemic [5–8]. Nevertheless, the majority of the evidence on the cause-effect relationship or predictive correlation between online information and vaccine hesitancy seem to support a positive correspondence between access to information content on social media (sometimes characterized by inaccuracies or even falsehoods) and hesitation to vaccinate [5, 9–14]. In particular, research carried out using representative samples and randomized control trials suggest that exposure to online misinformation has an effect on the unconditional acceptance of vaccines [5, 9]. These results are consistent with the majority of research on communication carried out since the 30s that have gradually consolidated the idea that mass com-

munication is not a silver bullet causing abrupt changes in the behavior of consumers/citizens, but rather an information flow that shapes opinions and values in the long term by providing an overall cognitive framing [15–18]. Moreover, another body of research based on online quasi-experiments suggests that the circulation of online misinformation is a good predictor of large-scale vaccine hesitancy [10–14]. Also in this case, the results seem to support established theories in communication studies by showing how the pollution of the public conversational arena ends up disrupting complex debates that require precise information and articulate critical thinking [19], as it is certainly the case for the vaccines debate. It is such a pollution of the informational quality of the debate that may set the conditions for a shift of certain segments of the public opinion towards hesitancy positions [12]. In some circumstances, even casual, brief exposure could cause significant perceptual and attitudinal changes towards vaccines [14]. In addition, the influence of extreme fringes of online no-vax communities seems to extend well beyond their boundaries, in terms of recruitment of new, more or less intentional, supporters for their views and behavioral prescriptions [13]. In conclusion, as proposed by Hansen and colleagues, media coverage appears to be a good predictor of vaccine adherence, and this implies that it is unlikely that infodemic waves will not condition the attitudes of the people who, without being prejudicial anti-vaxxers, are however confused and disoriented and could end up refusing to vaccinate as a precautionary response to the exposure to misleading content [11]. This state of informational incertitude associated to an overabundance of information circulating on a topic, including a significant fraction that designed to disorientate the audience, is indeed exactly the operational definition of an infodemic.

B. METHODOLOGICAL CONCERNS

From a methodological perspective, we think that the conclusions of the authors are based on i) data collection potentially biased and ii) misinterpretation of aggregated results. We expand on both points in the following.

Measuring the infodemic.— In one of the analyses reported in [1], the authors count the number of posts, specifically tweets in the case of Twitter, to track the infodemic. However, the infodemic phenomenon is far more complex than what is captured by just an increasing the volume of messages within an online social network: for instance, an online community affected by infodemic could maintain a relatively small volume of activity compared to a more healthy online community characterized by a higher level of social activity, among other distinctive factors [20]. The main difference between those two exemplary communities, concerning the infodemic phenomenon, concerns the quality and content of news shared, as well as the exposure to those news: sources pointing to web sites of conspiracy theo-

rists or fake news or highly biased political content, are more likely to sustain the infodemic even if their volume might be relatively low. In fact, measuring the infodemic level or the infodemic risk within an online community is a complex task [21–23], which requires to account for the semantic content of messages, not just their volume. Therefore, we recommend a more critical inspection of the available literature and to use more reliable infodemic indicators. More specifically, we refer to [22, 24–27] for recent literature about additional grounded methods for quantitatively describing the infodemic phenomenon.

Counterfactual arguments.— The analyses carried out by Valensise et al [1] are of a purely descriptive nature and mostly focused on the description of the events surrounding the suspension of the Astrazeneca vaccine.

For instance, one of the central results in [1] concerns the sudden drop, limited to about 3-4% , in the acceptance rate following the EMA declarations about the Astrazeneca vaccine. However, a definitive analysis cannot rely on the observation of a specific short-time effect: in social systems, public health announcements and actions might have mid- and long-term effects which are not analyzed, and therefore confirmed or excluded, in [1]. Therefore, the observation reported by the authors is not conclusive and requires further analysis across multiple time scales.

Nevertheless, the conclusions drawn by the authors – when extrapolated beyond the boundaries of their working hypothesis – could be misleading, since a causal reading of the result is presented: namely, the absence of infodemic effects on vaccine hesitation. Over the years several scholars have been engaged in creating tools and questioning the correct interpretation of different statistical methods [28–30]. From this point of view it is therefore desirable that in the field of policy-oriented studies that suggest implications of a causal nature there would be a greater care and adherence to the correct communication of scientific information [28].

Selection bias.— The results reported in [1] are focused on only 6 countries, likely due to limited availability of social network data for a larger set of countries. However, it can be observed a high variability in the slopes of the trend used to evaluate the stability of vaccine hesitancy in the sampled population across a variety of countries. Moreover, the countries chosen for comparison against social media data present a marked homogeneity in the pandemic trend and in other important variables, such as the location within a homogeneous cultural area. From this point of view, classic comparative studies that have encountered the same problem suggest instead to maximize heterogeneity by taking as cases countries that present markedly distant characteristics in the pandemic trend or by cultural hummus [31, 32]. For example, in the same time period analyzed, South American countries or India presented pandemic trends very different from the countries considered and simultane-

ously would have maximized cultural differences. The comparative strategies are in fact used when it is impossible to obtain a statistical inference but there is the desire to outline a picture that identifies possible factors underlying cause-effect relationships.

Data set completeness: Twitter.— According to [1], Twitter posts have been obtained from searching the full historic archive, by means of the new API provided by Twitter. However, we argue that this data set might be very biased, since the service does not provide access to deleted tweets or tweets posted by deleted accounts during the observational period. To test this hypothesis, we have accessed the data collected by the COVID-19 Infodemic Observatory, including over 2 billions posts between 22th Jan 2020 and 17th July 2021, restricting our attention on the same observational period used in [1] and in an overlapping pool countries, namely Denmark (DNK), France (FRA), Germany (DEU), Italy (ITA), Spain (ESP) and United Kingdom (GBR). We have randomly selected 20,000 tweets per country, excluding retweets and equally distributing the number of tweets pointing to unreliable and reliable sources (we refer to [22] for details about the classification of sources). Results, reported in Tab. I, show that – regardless of source classification – between 30% and 53% of tweets are no more accessible: this simple fact might have dramatic effects on naive infodemic measures, such as the one based on simple counting.

Country	Reliable (sample size)	% still available	Unreliable (sample size)	% still available
DEU	10,000	49.1%	10,000	44.7%
DNK	9,553	42.2%	3,953	30.9%
ESP	10,000	50.3%	10,000	47.4%
FRA	10,000	52.6%	10,000	53.1%
GBR	10,000	48.8%	10,000	47.3%
ITA	10,000	49.6%	10,000	46.1%

TABLE I. Number of randomly sampled tweets (excluding retweets) from six countries. (For Denmark we include all original messages recorded.) We report the number of tweets pointing to reliable sources, the number of tweets pointing to potentially unreliable sources, together with the corresponding fraction of tweets that can be successfully retrieved from Twitter as of 19th July 2021, i.e., the ones not deleted by the users or from banned users after the Twitter campaign against misinformation started^a.

Data set completeness: Vaccine hesitancy.— In Appendix A and B, we perform a set of analyses to better understand to which extent the results reported in [1] depend on the completeness and the quality of data used for vaccine hesitancy, as well as on the quality of infodemic indices.

Specifically, we have performed an analysis similar in spirit to the one reported in [1], by performing non-parametric tests on the data provided by [1] and infodemic indicators publicly available from the COVID-19 Infodemic Observatory[33] and point out a possible, although weak, correlation: the result of the bootstrap

FIG. 1. Correlation analysis, per country, comparing the evolution of vaccine hesitancy – obtained from the MIT COVID-19 Survey – and the evolution of the infodemic risk index – obtained from the COVID-19 Infodemic Observatory. The observational period is between 6th July 2020 and 28th March 2021. We observe both positive (colored in red) and negative (colored in light blue) correlations, as well as the lack of correlations (colored in light grey), for distinct countries, highlighting that global and simplistic analyses are not sufficient to assess the interplay between vaccine hesitancy and infodemic spreading. See Tab. III and App. B for details. Note that the label *all* refers to the overall analysis performed without stratifying by country: in this case the correlation is non significant, denoting a striking difference between country-stratified results and the aggregated one, reminiscent of the Simpson’s paradox.

analysis for this small sample, gives a p -value of 0.116, suggesting the presence of a mild effect, rather than the lack of it. Furthermore, we obtain significant correlations (p -values smaller than 0.01) when we perform the same analysis using subsets of countries with lower average daily responses. See Fig. 7 for details.

However, even this result would not be conclusive to make any assessment about the (causal) relationship between infodemic spreading and vaccine hesitancy, since an independent data source for vaccine hesitancy is still needed to validate the result. To this aim, we have considered 19 survey waves from the MIT COVID-19 Survey (see App. B for details) and performed a correlation analysis against the evolution of the infodemic risk index [22], associated to the topic of the COVID19 epidemic in general, in the same observational period, either stratified by country or overall. A summary of the results is shown in Fig. 1, highlighting the existence of a broad spectrum of patterns including both negative and positive correlations, as well as no correlations at all. Our findings suggest that the interplay between infodemic spreading and vaccine hesitancy is rather complex and requires more sophisticated analysis to be understood.

We have also performed an additional analysis, using an independent and more complete data set concerning messages posted to Twitter. We consider tweets in the Twitter Firehose on COVID-19 (see App. A for details)

FIG. 2. Correlation analysis, per country, comparing the evolution of vaccine hesitancy – obtained from the MIT COVID-19 Survey – and the evolution of a trivial proxy for infodemic, namely the volume of tweets about vaccines – obtained from the Twitter Firehose on COVID-19. The overlapping observational period is between the 5th and 19th waves of the MIT survey data set, namely between 1st September 2020 and 28th March 2021. We observe both positive (colored in red) and negative (colored in light blue) correlations, as well as the lack of correlations (colored in light grey), for distinct countries, highlighting that global and simplistic analyses are not sufficient to assess the interplay between vaccine hesitancy and infodemic spreading. See Tab. II and App. A for details. Note that the label *all* refers to the overall analysis performed without stratifying by country: in this case the correlation is non significant, denoting a striking difference between country-stratified results and the aggregated one, reminiscent of the Simpson’s paradox.

and consider, as in [1], the volume of geotagged tweets about vaccines as a proxy for the size of the ongoing infodemic. Similarly to the previous analysis, we compare the evolution over time of such cumulative volumes per country, and overall, against the evolution of vaccine hesitancy measured from the MIT COVID-19 Survey. Results, shown in Fig. 2 and summarized in Tab. II confirm the previous analysis, where the infodemic risk index has been used, further confirming that the interplay between infodemic spreading and vaccine hesitancy requires more sophisticated analysis to be understood.

C. CONCLUSIONS

In the case of severe health crises such as the current pandemic, non-pharmacological interventions, and more specifically behavioral prescriptions, may have an impact that may be comparable to that of pharmacological ones, insofar as such interventions affect in a systematic way people’s level of compliance to public health recommendations [34]. For this reason, extreme care should be used toward the process of validation and diffusion of scientific evidence that may impinge upon the public opinion as well as upon policymakers’ orientations. If pharma-

cological interventions can only be released after a long and extremely selective process of screening and evaluation, an analogous level of screening, to the extent to which this is feasible and reasonable, should be exerted on claims that may have large scale socio-behavioral effects with major health consequences if accepted as true. Whether social media have an impact on health-related behaviors in the middle of a pandemic is an issue of burning importance. In assessing its effects, we need to apply the highest methodological standards, also taking into account that strong claims in this regard might easily be appropriated and tweaked for ideological reasons.

Even if vaccination is clearly a pharmaceutical intervention, any kind of action that is going to influence people's motivation to vaccinate is clearly a potential component of a non-pharmaceutical, critical counterpart of the former. These days, the media are publishing interviews of physicians treating seriously ill, non-vaccinated COVID-19 patients who, when asked why they did not want to receive the vaccine, report that they were convinced by information and opinion coming from different sources, among which social media play an outstanding role [35]. This is clearly anecdotal evidence, but it gives us an extra stimulus in adopting the strictest possible standards of testing for critically sensitive claims. Social media have become a deeply ingrained presence in everyday life, and they have existed for barely a couple of decades. We know very little about their long-term socio-cognitive and behavioral effects, and we probably do not have yet the appropriate scientific toolbox for this kind of analysis [36]. We are literally experimenting in the wild. We need to be cautious accordingly.

Not only we know little about the deep structural forces at work. Often our own access to data is critically limited. As a norm, no matter how large the samples we are able to build in absolute terms, our sampling of social media platforms is necessarily partial and subject to many biases, of which we must be constantly aware. For instance, the interpretation of results from Twitter is typically very difficult and many quantities have been introduced in the past. Tweets volume alone has already been proven not to be particularly relevant for the analysis of societal trends (see for example [37] and references therein for a discussion on election forecasting, as well as [38] for the limits to predictability in partially observed systems). Two different approaches are

then available: either considering semantic content analysis (e.g., good/bad news content, message, tweet,...), or focusing on topology, where the sign of the message is determined by the community of users one belongs to [39–42]. More specifically, any infodemic analysis should account for the content and/or type of news shared and the user's exposure to them [22, 23, 43–46]. Unfortunately, none of them has been pursued in the case of [1]. Moreover, if, as in the present case, data analysis is carried out on a very partial data set, this might in principle bias the results severely, unless finite-size effects are dealt with properly [47], or a statistical reconstruction [48] of the underlying social network is attempted. Finally, it is worth mentioning that the problem of inferring causal relationships between phenomena co-occurring in the physical and digital spaces is usually challenging, and requires an appropriate experimental setup for gathering data suitable for a meaningful analysis of the cause/effect relationships. Even in ideal setups, there can be a variety of factors leading to a lack of observation which, however, does not easily allow one to assess a lack of effect: this might be the case of small sample data, noisy data, statistically biased analyses, low resolution power (i.e., the complementary of Type II error), so forth and so on.

What is the actual impact of social media on vaccine hesitancy remains an open question, but in the light of the above considerations, we think that applying a precautionary principle is the most prudent conduct, assuming that they could have a potentially negative effect unless otherwise proven. And this means, in particular, that proof of their irrelevance should be especially compelling and methodologically rigorous. We do not think the paper we are commenting upon provides us with such a level of evidence, and we urge media and policymakers not to underplay the potential threat posed by infodemics, and the consequences that an oversight in this regard could have on the public opinion, on individual and social behaviors, and ultimately on the efficacy of crucial non-pharmacological public health interventions.

ACKNOWLEDGMENTS

We thank the authors of [1] for sharing part of the processed data required to constructively cross-validate their results.

-
- [1] C. M. Valensise, M. Cinelli, M. Nadini, A. Galeazzi, A. Peruzzi, G. Etta, F. Zollo, A. Baronchelli, and W. Quattrociocchi, *The covid-19 infodemic does not affect vaccine acceptance* (2021), arXiv:2107.07946.
 - [2] R. M. Jacobson, J. L. S. Sauver, and L. J. F. Rutten, in *Mayo Clinic Proceedings* (Elsevier, 2015), vol. 90, pp. 1562–1568.
 - [3] N. E. MacDonald et al., *Vaccine* **33**, 4161 (2015).
 - [4] D. A. Salmon, M. Z. Dudley, J. M. Glanz, and S. B. Omer, *Vaccine* **33**, D66 (2015).
 - [5] S. Loomba, A. de Figueiredo, S. J. Piatek, K. de Graaf, and H. J. Larson, *Nature human behaviour* **5**, 337 (2021).
 - [6] N. Puri, E. A. Coomes, H. Haghbayan, and K. Gunaratne, *Human vaccines & immunotherapeutics* **16**, 2586 (2020).
 - [7] E. Karafillakis, C. Simas, C. Jarrett, P. Verger, P. Peretti-Watel, F. Dib, S. De Angelis, J. Takacs, K. A. Ali, L. P. Celentano, et al., *Human vaccines & im-*

- munotherapeutics (2019).
- [8] J. Roozenbeek, C. R. Schneider, S. Dryhurst, J. Kerr, A. L. Freeman, G. Recchia, A. M. Van Der Bles, and S. Van Der Linden, *Royal Society open science* **7**, 201199 (2020).
 - [9] D. Freeman, B. S. Loe, L.-M. Yu, J. Freeman, A. Chadwick, C. Vaccari, M. Shanyinde, V. Harris, F. Waite, L. Rosebrock, et al., *The Lancet Public Health* **6**, e416 (2021).
 - [10] D. Allington, S. McAndrew, V. Moxham-Hall, and B. Duffy, *Psychological medicine* pp. 1–12 (2021).
 - [11] P. R. Hansen and M. Schmidtblaicher, *Journal of business & economic statistics* **39**, 259 (2021).
 - [12] S. L. Wilson and C. Wiysonge, *BMJ Global Health* **5**, e004206 (2020).
 - [13] V. Carrieri, L. Madio, and F. Principe, *Health economics* **28**, 1377 (2019).
 - [14] C. Betsch, F. Renkewitz, T. Betsch, and C. Ulshöfer, *Journal of health psychology* **15**, 446 (2010).
 - [15] C. Fuchs and J. L. Qiu, *Ferments in the field: Introductory reflections on the past, present and future of communication studies* (2018).
 - [16] R. L. Holbert, R. K. Garrett, and L. S. Gleason, *Journal of communication* **60**, 15 (2010).
 - [17] E. Katz, *The Public Opinion Quarterly* **51**, S25 (1987).
 - [18] H. Blumer (1933).
 - [19] S. Hilgartner and C. L. Bosk, *American journal of Sociology* **94**, 53 (1988).
 - [20] V. Tangcharoensathien, N. Calleja, T. Nguyen, T. Purnat, M. D’Agostino, S. Garcia-Saiso, M. Landry, A. Rashidian, C. Hamilton, A. AbdAllah, et al., *Journal of Medical Internet Research* **22**, e19659 (2020), URL <https://doi.org/10.2196/19659>.
 - [21] M. Luengo-Oroz, K. H. Pham, J. Bullock, R. Kirkpatrick, A. Luccioni, S. Rubel, C. Wachholz, M. Chakchouk, P. Biggs, T. Nguyen, et al., *Nature Machine Intelligence* **2**, 295 (2020).
 - [22] R. Gallotti, F. Valle, N. Castaldo, P. Sacco, and M. D. Domenico, *Nature Human Behaviour* **4**, 1285 (2020), URL <https://doi.org/10.1038/s41562-020-00994-6>.
 - [23] K.-C. Yang, F. Pierri, P.-M. Hui, D. Axelrod, C. Torres-Lugo, J. Bryden, and F. Menczer, *Big Data & Society* **8**, 205395172110138 (2021), URL <https://doi.org/10.1177/20539517211013861>.
 - [24] A. Gruz, M. D. Domenico, P. L. Sacco, and S. Briand, *Big Data & Society* **8**, 205395172110211 (2021), URL <https://doi.org/10.1177/20539517211021115>.
 - [25] E. Bonnevie, J. Sittig, and J. Smyser, *Big Data & Society* **8**, 20539517211013867 (2021).
 - [26] K. T. Gradoń, J. A. Hołyst, W. R. Moy, J. Sienkiewicz, and K. Suchecki, *Big Data & Society* **8**, 20539517211013848 (2021).
 - [27] M. Green, E. Musi, F. Rowe, D. Charles, F. D. Pollock, C. Kypridemos, A. Morse, P. Rossini, J. Tulloch, A. Davies, et al., *Big Data & Society* **8**, 20539517211013869 (2021).
 - [28] S. Athey, *Science* **355**, 483 (2017).
 - [29] J. D. Angrist and J.-S. Pischke, *Mastering metrics: The path from cause to effect* (Princeton university press, 2014).
 - [30] K. Imai, G. King, E. A. Stuart, et al., *Journal of the Royal Statistical Society Series A* **171**, 481 (2008).
 - [31] N. J. Smelser, *Comparative methods in the social sciences* (Quid Pro Books, 2013).
 - [32] G. Sartori, *Journal of theoretical politics* **3**, 243 (1991).
 - [33] Note1, <https://covid19obs.fbk.eu/#/>.
 - [34] N. Perra, *Physics Reports* (2021).
 - [35] Note2, see for instance <https://www.al.com/news/2021/07/im-sorry-but-its-too-late-alabama-doctor-on-treating-unvaccinated-dying-covid-patients.html>.
 - [36] N. Bartley, A. Abeliuk, E. Ferrara, and K. Lerman, in *13th ACM Web Science Conference 2021* (ACM, 2021), URL <https://doi.org/10.1145/3447535.3462491>.
 - [37] G. Caldarelli, A. Chessa, F. Pammolli, G. Pompa, M. Puliga, M. Riccaboni, and G. Riotta, *PloS one* **9**, e95809 (2014).
 - [38] A. Abeliuk, Z. Huang, E. Ferrara, and K. Lerman, *Scientific Reports* **10** (2020), URL <https://doi.org/10.1038/s41598-020-77091-1>.
 - [39] M. Stella, M. Cristoforetti, and M. D. Domenico, *PLOS ONE* **14**, e0214210 (2019), URL <https://doi.org/10.1371/journal.pone.0214210>.
 - [40] O. Artime, V. d’Andrea, R. Gallotti, P. L. Sacco, and M. D. Domenico, *Scientific Reports* **10** (2020), URL <https://doi.org/10.1038/s41598-020-71231-3>.
 - [41] G. Caldarelli, R. De Nicola, M. Petrocchi, M. Pratelli, and F. Saracco, *EPJ Data Science* **10**, 34 (2021).
 - [42] Sacco, Gallotti, Pilati, Castaldo, and D. Domenico, To appear in *Social Science & Medicine* (2021).
 - [43] C. Shao, P.-M. Hui, L. Wang, X. Jiang, A. Flammini, F. Menczer, and G. L. Ciampaglia, *PLOS ONE* **13**, e0196087 (2018), URL <https://doi.org/10.1371/journal.pone.0196087>.
 - [44] C. Shao, G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer, *Nature Communications* **9** (2018), URL <https://doi.org/10.1038/s41467-018-06930-7>.
 - [45] E. Ferrara, S. Cresci, and L. Luceri, *Journal of Computational Social Science* **3**, 271 (2020), URL <https://doi.org/10.1007/s42001-020-00094-5>.
 - [46] S. González-Bailón and M. D. Domenico, *Proceedings of the National Academy of Sciences* **118**, e2013443118 (2021), URL <https://doi.org/10.1073/pnas.2013443118>.
 - [47] M. Serafino, G. Cimini, A. Maritan, A. Rinaldo, S. Suweis, J. R. Banavar, and G. Caldarelli, *Proceedings of the National Academy of Sciences* **118** (2021).
 - [48] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli, *Nature Reviews Physics* **1**, 58 (2019).
 - [49] Note3, for details, see <https://developer.twitter.com/en/docs/labs/covid19-stream/filtering-rules>.
 - [50] Note4, <https://covidsurvey.mit.edu/>.

Appendix A: Results from the Twitter Firehose on COVID-19

Overview of social media data.— We use tweets collected by means of the special Twitter’s endpoint dedicated to COVID-19 research. Specifically, between 1st July 2020 and 15th July 2021, we have gathered, in real time, more than 2 billions messages publicly posted worldwide to Twitter and containing one or more keywords or hashtags among the 500+ terms manually annotated by humans to filter posts related to the public discussion about COVID-19. Note that this data set is not affected by *a posteriori* actions, such as user ban or tweet deletion, thus providing a unique and suitable data set for advanced analysis. In this appendix, we focus on data collected between the 1st September 2020 and the 15th July 2021 [49].

In Fig. 3 we report the terms captured by the filter of the Twitter Firehose on COVID-19 that we use for the present analysis, including 18 among the most representative languages used on Twitter. Specifically, we focused on terms related to vaccination, anti-vax campaigns and the names of known vaccines to date. Additionally, we consider only the small fraction (about 1%) of geotagged tweets, to guarantee accurate information about their location. The overall number of tweets surviving our filter is of about 11M: the evolution of the cumulative number of posts, stratified by vaccine type, is shown in Fig. 4. Figures 5 and 6 show, respectively, the evolution of the number of posts and the number of unique users stratified by vaccine type and country, for the six countries discussed in [1].

Overview of vaccine hesitancy data.— We use two distinct data sets:

- The slopes of vaccine acceptance rate provided by [1], as obtained from the COVID-19 Symptom Survey, part of the Facebook Data for Good project;
- The evolution, in 19 waves, of vaccine acceptance over time provided by the MIT COVID-19 Survey[50], covering the period between 6th July 2020 and 28th March 2021. This data set has been built by MIT in collaboration with Facebook, Johns Hopkins University (JHU), the World Health Organization (WHO), and the Global Outbreak Alert and Response Network (GOARN).

Analyzing MIT COVID-19 Survey versus Twitter Firehose on COVID-19.— In this section, we consider the observational period overlapping between the MIT COVID-19 Survey and our Twitter Firehose on COVID-19, to perform an analysis similar in spirit to the one proposed in [1], where the volume of posts about the public discussion around vaccines is used, naively, as a proxy for the intensity of the infodemic phenomenon.

Specifically, we consider the whole temporal evolution of the percentage of respondents who answered negatively

to the question about vaccine acceptance, while considering the evolution of the number of tweets in the same period. For the latter, the measure used for the analysis corresponds to the average amount measured for that country during a specific wave. The overlapping period between the two data sets corresponds to waves 5–19 in the MIT COVID-19 Survey. The significance is evaluated with a bootstrap approach, by reshuffling the two time series for 10,000 times and by comparing the observed value of Spearman’s ρ with the distribution of the correlation values obtained under the null hypothesis. The size of the statistical test is fixed, *a priori*, to 10%, corresponding to a 90% significance. Results are reported in Tab. II and visualized in Fig. 2, showing that distinct countries exhibit distinct levels of correlation between the infodemic proxy and vaccine hesitancy, with a rich repertoire of scenarios, ranging from positive correlations (i.e., vaccine hesitancy *increases* with country’s infodemic risk) to negative correlations (i.e., vaccine hesitancy *decreases* with country’s infodemic risk). We conclude that simplistic analyses are not enough to assess the complex interplay between these two social and behavioral phenomena, and more sophisticated analyses are needed to account for different socio-cultural, behavioral, environmental and epidemiological factors.

Country	Corr	P-value	Effect
TUR	0.54	0.02	Positive Effect
IND	0.52	0.02	Positive Effect
EGY	0.43	0.05	Positive Effect
BGD	0.40	0.07	Positive Effect
PHL	0.21	0.22	No Effect
ARG	0.16	0.28	No Effect
BRA	0.14	0.30	No Effect
COL	-0.08	0.38	No Effect
NGA	-0.13	0.32	No Effect
MYS	-0.16	0.28	No Effect
THA	-0.28	0.15	No Effect
FRA	-0.29	0.14	No Effect
IDN	-0.31	0.12	No Effect
<i>all</i>	<i>-0.31</i>	<i>0.13</i>	<i>No Effect</i>
JPN	-0.35	0.10	Negative Effect
PAK	-0.35	0.10	Negative Effect
GBR	-0.45	0.05	Negative Effect
VNM	-0.47	0.04	Negative Effect
ITA	-0.52	0.03	Negative Effect
ROU	-0.57	0.02	Negative Effect
MEX	-0.68	0.01	Negative Effect
USA	-0.70	< 0.01	Negative Effect
DEU	-0.75	< 0.01	Negative Effect
POL	-0.78	< 0.01	Negative Effect

TABLE II. Spearman’s correlation analysis between vaccine hesitancy evolution obtained from the MIT COVID-19 Survey and the volume of messages posted to Twitter obtained from the Twitter Firehose on COVID-19. A test size of 10% (corresponding to 90% significance) is used to assess the presence or absence of an effect.

Language	Subjects		Verbs	Participes	Root	Anti-vax
English	vaccine(s)	vaccination(s)	vaccine	vaccinated	vaccin	anti-vax
Italian	vaccino(i)	vaccinazione(i)	vaccinare	vaccinato	vaccin	no-vax
Spanish	vacuna(s)	vacunaci3n/vacunas	vacunar	vacunado	vacun	anti-vacunas
Catalan	vacuna(s)	vacunaci3/vacunes	vacunar	vacunat	vacun	
Portuguese	vacina(s)	vacina33o/3es	vacinar	vacinado	vacin	anti-vacinas
French	vaccin (s)	vaccination(s)	vacciner	vaccin3	vaccin	anti-vaccins
German	Impfstoff (e)	Impfung(en)	impfen	geimpft/ geimpften	impf	anti-impf
Russian	вакцина	Вакцинация/прививки	вакцинировать	вакцинированный	вакцина	антивакс
Polish	szczepionka/i	szczepienie	szczepi3	szczepione	szczepi	Anty-szczepionki
Danish	vaccine (r)	vaccination (er)	vaccinere	vaccineret	vaccin	
Swedish	vaccin(er)	vaccinering	vaccinera	vaccinerad	vacciner	Vaccinationsmotst3nd
Turkish	aşı(lar)	aşılama	aşılamak	aşılanmış	aşı	aşı karşıtı
Indonesian	vaksin	vaksinasi	divaksinasi	divaksinasi	vaksin	
Japanese	ワクチン	予防接種	予防接種	予防接種		
Thai	วักซึน	การฉีดวักซึน	ฉีดวักซึน	ฉีดวักซึน		
Korean	백신	백신 접종	예방 접종	예방 접종		안티 백신
Arabic		مصل	تلقيح	لقح	تطعيم	
Hindi	टीका (टीके)	टीका (टीकाकरण)	टीका लगाना	टीका		रंटी टीका लगाना

FIG. 3. Terms, from 18 most representative languages on Twitter, used to filter the Twitter Firehose on COVID-19 for the present analysis. The terms are related to vaccination. Additionally, we considered also the following terms related to vaccine names: astrazeneca, oderna, johnson & johnson, sputnik, pfizer, biontech.

FIG. 4. Cumulative number of messages posted to Twitter over time, concerning different vaccines (Pfizer, Moderna, Astrazeneca, Sputnik V, Johnson & Johnson are considered). Only geotagged posts in the Twitter Firehose on COVID-19 are considered, to guarantee high quality of data selection. Labels refer to manually inspected events about the massive COVID-19 vaccination campaign, from announcements of efficacy to precautionary suspension of the Astrazeneca vaccine.

Appendix B: Results from the COVID-19 Infodemic Observatory

Overview of social media data.— From 23rd January 2020 to 15th July 2021, we collected from the Twitter Filter API messages containing terms associated with the COVID-19 pandemic from a medical per-

spective (namely: coronavirus, ncov, Wuhan, covid19, covid-19, sarscov2, covid). We collected the totality of tweets associated with these words up until 5th February 2020, then the fraction of tweets included in our filter passed the 1% of the total, hitting the limit set by Twitter for basic accounts. To associate tweets with countries, we use the information about the user location included in the information provided by the Twitter API.

FIG. 5. Cumulative number of messages posted to Twitter over time in six distinct countries, concerning different vaccines (Pfizer, Moderna, AstraZeneca and Johnson & Johnson are considered). Only geotagged posts in the Twitter Firehose on COVID-19 are considered, to guarantee high quality of data selection.

FIG. 6. Cumulative number of unique users posting to Twitter over time, discussing about different vaccines (Pfizer, Moderna, AstraZeneca and Johnson & Johnson are considered). Only geotagged posts in the Twitter Firehose on COVID-19 are considered, to guarantee high quality of data selection.

To evaluate the amount of unreliable news circulating every day in each country, we first selected all tweets associated with that country containing an URL, and then checked whether the associated web domain found a match with a database we created by integrating multiple publicly available databases of manually-checked web

domains (see reference [22] for more details). Using the database, we classify news in one of seven categories: (Science, Mainstream Media, Satire, Clickbait, Political, Fake/Hoax, Conspiracy/Junk Science) plus two types of unclassified news (unknown, when not in the database, or shadow, when the link is compressed by non-vanity

URL shorteners that we could not access at the time of the data pre-elaboration). News belonging to Science and Mainstream Media have been considered as reliable, whereas news categorized as Satire, Clickbait, Political, Fake/Hoax or Conspiracy/Junk Science have been considered unreliable.

As extensively shown in [22], the data sample is not biased with respect to the type of analysis and conclusions about infodemic risk analysis.

Analyzing Facebook Data for Good versus COVID-19 Infodemic Observatory.— In this section, we consider the observational period overlapping between the COVID-19 Symptom Survey and the COVID-19 Infodemic Observatory, to perform a correlation analysis similar in spirit to the one reported in [1]. Specifically, as infodemic indices we consider, for each country, the unreliability index of the infodemic observatory as well as the ratio between the unreliability index for the vaccines firehose data divided by the unreliability index of the infodemic observatory, to capture the fraction of misinformation that is specifically related to vaccines. We perform the non-parametric Spearman correlation analysis and compare the result against $n = 10000$ bootstrapped surrogate data, under the null hypothesis of no correlation between the two aforementioned measures. Therefore, we evaluate the corresponding p -value and report results in Fig. 7. The analysis shows that by selecting countries with different coverage, here estimated in terms of average number of messages related to vaccines posted, we find the presence of signals with decreasing p -value. From this analysis, we learn that i) country selection plays an important role in assessing the presence of a signal, and ii) that signal strength might depend on the coverage of social activities. Therefore, without a clear understanding of coverage, it is dangerous to assess the presence or the lack of any signal.

Analyzing MIT COVID-19 Survey versus COVID-19 Infodemic Observatory.— In this section, we consider the observational period overlapping between the MIT COVID-19 Survey and the COVID-19 Infodemic Observatory, to perform a deeper analysis than the previous one. Specifically, we consider the whole temporal evolution of the percentage of respondents who answered negatively to the question about vaccine acceptance, while considering the evolution of the infodemic risk index (IRI) in the same period. For the latter, the measure used for the analysis corresponds to the average IRI for that country during a specific wave. Similarly to the previous analysis, the significance is evaluated with a bootstrap approach, by reshuffling the two time series

for 10,000 times and by comparing the observed value of Spearman’s ρ with the distribution of the correlation values obtained under the null hypothesis. The size of the statistical test is fixed, *a priori*, to 10%, corresponding to a 90% significance. Results are reported in Tab. III and visualized in Fig. 1, showing that distinct countries exhibit distinct levels of correlation between infodemic risk and vaccine hesitancy, with a rich repertoire of scenarios, ranging from positive correlations (i.e., vaccine hesitancy *increases* with country’s infodemic risk) to negative correlations (i.e., vaccine hesitancy *decreases* with country’s infodemic risk). We conclude here again that simplistic analyses are not enough to assess the complex interplay between these two social and behavioral phenomena, and more sophisticated analyses are needed to account for different socio-cultural, behavioral, environmental and epidemiological factors.

Country	Corr	P-value	Effect
ROU	0.69	< 0.01	Positive Effect
DEU	0.62	< 0.01	Positive Effect
USA	0.50	0.02	Positive Effect
BRA	0.46	0.03	Positive Effect
MEX	0.45	0.03	Positive Effect
GBR	0.31	0.09	Positive Effect
IDN	0.26	0.13	No Effect
JPN	0.09	0.35	No Effect
MYS	0.07	0.39	No Effect
ITA	0.03	0.45	No Effect
VNM	0.02	0.47	No Effect
ARG	0.02	0.48	No Effect
PAK	-0.05	0.42	No Effect
THA	-0.08	0.36	No Effect
NGA	-0.16	0.24	No Effect
BGD	-0.17	0.23	No Effect
POL	-0.21	0.20	No Effect
PHL	-0.23	0.17	No Effect
<i>all</i>	<i>-0.24</i>	<i>0.15</i>	<i>No Effect</i>
TUR	-0.28	0.12	No Effect
IND	-0.57	0.01	Negative Effect
COL	-0.58	0.01	Negative Effect
FRA	-0.64	< 0.01	Negative Effect
EGY	-0.65	< 0.01	Negative Effect

TABLE III. Spearman’s correlation analysis between vaccine hesitancy evolution obtained from the MIT COVID-19 Survey and the infodemic risk index obtained from the COVID-19 Infodemic Observatory. A test size of 10% (corresponding to 90% significance) is used to assess the presence or absence of an effect.

FIG. 7. Spearman's rank correlation analysis and bootstrap results between drop in vaccine acceptance (from [1]) and infodemic indicators (from COVID-19 Infodemic Observatory^a). **Left:** all countries in the data set, here we use as infodemic index the ratio of unreliable tweets related to vaccines, divided by the ratio of unreliable tweets related to COVID-19; **right:** subset of countries with number of tweets related to vaccines in the bottom 50% of the distribution. Here we use, as infodemic index, the fraction of unreliable tweets related to COVID-19.