

Comment

# Opportunities and challenges of using social media big data to assess mental health consequences of the COVID-19 crisis and future major events

Martin Tušl<sup>1</sup> · Anja Thelen<sup>2</sup> · Kailing Marcus<sup>3</sup> · Alexandra Peters<sup>4</sup> · Evgeniya Shalaeva<sup>5</sup> · Benjamin Scheckel<sup>2</sup> · Martin Sykora<sup>6</sup> · Suzanne Elayan<sup>6</sup> · John A. Naslund<sup>7</sup> · Ketan Shankardass<sup>8</sup> · Stephen J. Mooney<sup>9</sup> · Marta Fadda<sup>10</sup> · Oliver Gruebner<sup>11,12</sup>

Received: 16 May 2022 / Accepted: 17 June 2022

Published online: 27 June 2022

© The Author(s) 2022 **OPEN**

## Abstract

The present commentary discusses how social media big data could be used in mental health research to assess the impact of major global crises such as the COVID-19 pandemic. We first provide a brief overview of the COVID-19 situation and the challenges associated with the assessment of its global impact on mental health using conventional methods. We then propose social media big data as a possible unconventional data source, provide illustrative examples of previous studies, and discuss the advantages and challenges associated with their use for mental health research. We conclude that social media big data represent a valuable resource for mental health research, however, several methodological limitations and ethical concerns need to be addressed to ensure safe use.

**Keywords** Social media · Big data · Mental health · COVID-19

## 1 Introduction

The COVID-19 pandemic has caused detrimental global consequences for public health, healthcare systems, economies, and society in general. It has disrupted day-to-day activities and exposed deep societal inequities as the impacts of the pandemic have been particularly severe among people from socio-economically disadvantaged backgrounds and other marginalized groups [1, 2]. Measures and restrictions including social distancing, closures of non-essential businesses, schools, and national borders were implemented around the world to counter the spread of the virus. While these measures were aimed at “flattening the curve”, they resulted in increases in the risk of adverse mental health consequences [3]. Phenomena such as social isolation, movement restrictions, and economic hardships have significantly contributed

✉ Martin Tušl, martin.tusl@uzh.ch | <sup>1</sup>Public and Organizational Health, Epidemiology, Biostatistics, and Prevention Institute, Center of Salutogenesis, University of Zurich, Zurich, Switzerland. <sup>2</sup>Institute for Health Economics and Clinical Epidemiology, The University Hospital of Cologne, University of Cologne, Cologne, Germany. <sup>3</sup>Faculty of Medicine, University of Geneva, Geneva, Switzerland. <sup>4</sup>Infection Control Programme and WHO Collaborating Center on Patient Safety, University of Geneva, Geneva, Switzerland. <sup>5</sup>Division of Public Health Science, Westminster International University in Tashkent, Tashkent, Uzbekistan. <sup>6</sup>Centre for Information Management, Loughborough University, Loughborough, Leicestershire, UK. <sup>7</sup>Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, USA. <sup>8</sup>Department of Health Sciences, Wilfrid Laurier University, Waterloo, ON, Canada. <sup>9</sup>Department of Epidemiology, University of Washington, Seattle, WA, USA. <sup>10</sup>Faculty of Biomedical Sciences, University of Lugano, Lugano, Switzerland. <sup>11</sup>Department of Geography, University of Zurich, Zurich, Switzerland. <sup>12</sup>Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland.



to an increase in emotional distress [3, 4]. The pandemic's global impact on mental health is less immediate and visible than its physical consequences but may prove equally devastating.

Quantifying mental health consequences of a major global event such as the COVID-19 pandemic is extremely challenging. This is in part due to the cost and time associated with collecting and curating data related to mental health that can adequately represent experience on a broader scale. Systematic reviews and meta-analyses of national and multinational studies can shed light on the changes in mental health symptoms across different regions over the course of the pandemic [5, 6]. However, these methods have inferential challenges including inability to account for heterogeneity in the designs and methodologies of component studies, sampling methods, studied variables, and time-points of data collection [7]. Additionally, surveys often have limited coverage of under-represented groups of the population, such as those living in socio-economically disadvantaged settings, immigrants and refugees, individuals with mental health conditions, and ethnic minority groups [8].

In this commentary, we discuss the role of social media big data in overcoming the existing problems in mental health research. Social media big data can expand our understanding of mental health trajectories throughout the pandemic and across different population groups as they offer broader reach and ability to capture the nuanced social dynamics and enable access to a large amount of timely data "from the field". This is of utmost importance to identify and mitigate the immediate and long-term adverse psychological consequences associated with global crises such as the COVID-19 pandemic, and to understand and support the development of psychological resilience across populations.

## 2 Opportunities of using social media big data in mental health research

Social media are a digital-based technology that facilitates the sharing of ideas, thoughts, and information through the building of virtual networks and communities. They typically feature user-generated content and personalized profiles. The largest social media networks include Facebook, Instagram, Twitter, YouTube, and TikTok [9]. As of 2022, there were over 4.6 billion social media users globally [10]. The use of social media substantially increased during the pandemic as people had to find ways to fulfill their social needs when physical distancing was imposed [11]. Millions of people every day thus express their thoughts, moods, and emotions via social media platforms [10]. Some users even communicate openly about their everyday struggles with mental health through posts that include written expressions, images, and/or videos. The growing access to smartphones and resulting broader use of social media across the globe creates new opportunities to access data from population groups that may not typically participate in conventional research [12]. Social media data are available in real-time and can be accessed retrospectively, which mitigates the risk of recall bias [13] and offers opportunities for real-time surveillance that could inform preventive measures [14, 15]. They can be collected longitudinally, which facilitates investigation of causal relationships and making predictions about the development of mental health risks [16]. Furthermore, social media often allow to extrapolate information about the users' sociodemographic characteristics such as gender, age, nationality, or their social network, as well as the geographic location and temporal information [16, 17].

Due to the technological advancements in data processing, it is now possible to conduct accurate and time-efficient analysis of human emotions in big data. Machine learning and Natural Language Processing (NLP) techniques have been shown effective in decoding human emotions and have a great potential to unobtrusively help identifying mental health problems [18]. Machine learning is based on data-driven computational algorithms that can "learn" to recognize data patterns and make predictions. NLP, on the other hand, deals with computational processing of human natural language in text form, such as those found in social media. Advanced spatial epidemiological approaches combining tools such as EMOTIVE [19] or Stresscapes [20], that is, sentiment analysis of basic emotions or stress as found in social media, with spatiotemporal data analysis, allow for example, to track user-based information on the collective emotional state of a given community following the occurrence of significant events [17, 21]. These tools enable the detection of "hotspots" where high levels of negative emotions or stress may indicate a risk for mental health in reaction to local events [22]. Such methods could help to predict mental health development of users in specific regions based on their posts and reactions in real-time. Researchers could thus better understand the dynamics of crises such as COVID-19 and the consequences for mental health in specific subgroups and regions. Governmental and non-governmental organizations could also better adopt appropriate policies and measures, deliver targeted interventions, and allocate limited health care resources to the most affected regions and populations.

### 3 Social media big data and COVID-19

We list some examples of studies that used social media big data, mainly from Twitter, to assess the impact of the pandemic on mental health. A study [23] from the United States (US) analyzed 60 million tweets from the first wave of the pandemic (March–May 2020) and compared these with 40 million tweets from the same period in 2019. They used machine learning classifiers [24] to identify the social media language indicative of mental health problems (anxiety, depression, stress, and suicidal ideation), and found that overall, mental health symptomatic expressions increased by 14% during the COVID-19 period. The increase declined steadily over the period and eventually plateaued, possibly due to habituation effect or the effect of the imposed policy measures. A similar approach followed a study [25] that used machine learning models to identify expressions of stress, anxiety, loneliness, and the overall positive sentiment in a random sample of 5 million tweets/day from March to May 2020 and compared it to the same period in 2019. They identified a significant decrease in overall sentiment, and a significant increase in the expressions of stress, anxiety, and loneliness in 2020 compared to 2019.

Another US study [26] used machine learning and sentiment analysis techniques to assess the changes in social media use in 20 US cities and the shifts in public mood during the first months of the pandemic (January–April 2020). The number of tweets increased over the course of the pandemic and peaked during lockdowns suggesting that people may tend to use social media to cope with social isolation. Similar as in the previous study, there was an overall negative trajectory in the sentiment scores indicating a lower mood than prior to the pandemic. However, the sentiment analysis also revealed an increase in positivity of tweets, which the authors explain by “positivity bias” suggesting that the tweets may not reflect the underlying emotion of the population but rather the language used by media to discuss the topic. Users thus tend to post content that is more positive or optimistic than their true emotions.

Li and colleagues [27] used Twitter data to model spatiotemporal patterns of depressive symptoms associated with the COVID-19 pandemic in the US. They developed an algorithm that combines machine learning with Patient Health Questionnaire [28] lexicon to detect COVID-19 related depressive symptoms during the first months of the pandemic. They found a strong correlation between depressive symptoms and the number of increased COVID-19 cases.

Finally, a cohort study [29] aimed to develop machine learning models for predicting the change in depression and anxiety based on Google Search and YouTube data. They followed a cohort of undergraduate US students ( $N = 49$ ) from January to May 2020 and studied the changes in anxiety and depression scores in relation to data about the students' use of Google Search and YouTube. Deteriorating depression and anxiety conditions strongly correlated with behavioral changes in Google Search and YouTube use during the COVID-19 pandemic. The study illustrates a potential benefit of machine learning models based on social media big data for evaluating mental health conditions and their development over time.

### 4 Limitations and ethical concerns

Although social media provide an unprecedented access to a wealth of new data sources, their use for mental health research poses several challenges. First, it is challenging to validate social media measures against gold standard assessment tools. For instance, NLP algorithms scanning tweets may indicate the user is experiencing depression symptoms, but it is nearly impossible to assess whether the algorithm accurately reflects standard depression measures at scale and across cultures. Further, social media data restrict the study population to social media users, which may limit the representativeness of the sample and external validity of the investigations. Validity and reliability of the data can also be complicated by automated social media activity (e.g., “bot” accounts) and fake accounts, which can account for a large amount of social media posts [22]. Some of these limitations could be overcome by combining social media data sources with conventional survey-based methods that are prone to other biases (e.g., sampling error, selection bias). Choi and colleagues [30] have demonstrated a successful example of such social media data integration with traditional data collection for the forecasting of weekly suicide fatalities in the US. Finally, as illustrated in one of the studies [26], social media data are prone to various biases such as positivity bias, social desirability bias, or self-idealization which may create an inaccurate image of a user's mental health status and lead to false conclusions by researchers. Users may engage in some degree of self-idealization, however, research has shown that most social media users express their actual rather than their idealized identities [31].

Apart from methodological challenges, the use of social media big data for mental health research raises important ethical concerns. Although social media can be beneficial for mental health research purposes, the platforms themselves can negatively impact on users' mental health. Social media contribute to the spread of fake news and hate speech and frequent use of these platforms can lead to feelings of loneliness, depression, and anxiety [32]. Mental health research based on social media data also raises ethical concerns regarding the type of collected data, method of collection and analysis, publication, and any related privacy implications. Although there have been calls to set ethical standards for the use of social media in research [33], there are still no official guidelines in place [34]. Benton and colleagues [35] provide useful recommendations for social media health research regarding aspects such as informed consent, data protection, user interventions, and user identity protection. However, there is still a lack of comprehensive ethical standards in place for social media data use in mental health research. For instance, there remains debate on whether freely accessible social media data should be considered private or public [36]. Information that many would deem sensitive and/or private is therefore frequently shared in ways that was not intended by the user, with a potential for both government and corporate misuse [37]. Mental health research with social media data may involve user interventions such as identification of individuals potentially in need of psychological help. However, the users may have not intended their social media posts to be used for health research purposes, they never signed an informed consent to be part of the research, and they may perceive it as an inappropriate breach of privacy [35]. Due to the stigma around mental health problems, any identification of personal information can be harmful for the individual [37], especially when it involves social media users such as young children or older adults, for whom protection on social media may be insufficient [38].

## 5 Conclusion

Social media big data are a valuable resource for mental health research and may help us to act on acute issues in a timely manner which is of indisputable value for public mental health. Social media data could bring important insights about the scale, severity, and temporal changes of the global mental health burden stemming from the COVID-19 pandemic and future major events. In the long term, they could also contribute to the improvement of access to mental health services, and lead to a more efficient implementation of targeted prevention programs to decrease the incidence of mental disorders in specific areas and populations. Several methodological limitations and important ethical concerns need to be addressed to ensure safe use of such sensitive data.

**Acknowledgements** We acknowledge support by the Swiss School of Public Health (SSPH+) (to O.G.) through a mandate for the PhD course on Big Data in Public Health 2021. The manuscript is a direct outcome of this online seminar (SSPH+ PhD course website: <https://ssphplus.ch/en/graduate-campus/en/graduate-campus/course-program/>). M.T. received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 801076, through the SSPH+ Global PhD Fellowship Program in Public Health Sciences (GlobalP3HS) of the Swiss School of Public Health.

**Author contributions** M.T., A.T., K.M., A.P., E.S., contributed to the conceptualization of the manuscript, conducted literature review searches, and wrote the first draft of the manuscript under the supervision of O.G.; B.S., M.S., S.E., J.A.N., K.S., S.J.M., M.F., O.G. provided their expertise and contributed to the development of the final version of the manuscript. M.T. acts as the guarantor for this manuscript. All authors reviewed, read and approved the final manuscript.

**Funding** HORIZON EUROPE Marie Skłodowska-Curie Actions (801076).

### Declarations

**Competing interests** The authors declare that there are no competing interests.

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

## References

1. Wachtler B, et al. Socioeconomic inequalities and COVID-19—a review of the current international literature. *J Health Monit.* 2020;5:3–17. <https://doi.org/10.25646/7059>.
2. Ettman CK, et al. Low assets and financial stressors associated with higher depression during COVID-19 in a nationally representative sample of US adults. *J Epidemiol Community Health.* 2021;75:501–8. <https://doi.org/10.1136/jech-2020-215213>.
3. Brooks SK, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *Lancet.* 2020;395:912–20. [https://doi.org/10.1016/S0140-6736\(20\)30460-8](https://doi.org/10.1016/S0140-6736(20)30460-8).
4. Lakhan R, Agrawal A, Sharma M. Prevalence of depression, anxiety, and stress during COVID-19 pandemic. *J Neurosci Rural Pract.* 2020;11:519–25. <https://doi.org/10.1055/s-0040-1716442>.
5. Salazar de Pablo G, et al. Impact of coronavirus syndromes on physical and mental health of health care workers. *J Affect Disord.* 2020;275:48–57. <https://doi.org/10.1016/j.jad.2020.06.022>.
6. Ren X, et al. Mental health during the COVID-19 outbreak in China: a meta-analysis. *Psychiatr Q.* 2020;91:1033–45. <https://doi.org/10.1007/s11126-020-09796-5>.
7. Metelli S, Chaimani A. Challenges in meta-analyses with observational studies. *Evid Based Ment Health.* 2020;23:83–7. <https://doi.org/10.1136/ebmental-2019-300129>.
8. Laganà F, Elcherath G, Penic S, Kleiner B, Fasel N. National minorities and their representation in social surveys: which practices make a difference? *Qual Quant.* 2013;47:1287–314. <https://doi.org/10.1007/s11135-011-9591-1>.
9. Obar JA, Wildman SS. Social media definition and the governance challenge: an introduction to the special issue. *Telecommun Policy.* 2015;39:745–50. <https://doi.org/10.2139/ssrn.2663153>.
10. Kepios. Global Social Media Statistics. <https://datareportal.com/social-media-users>. 2022.
11. Feldkamp J. The rise of TikTok: the evolution of a social media platform during COVID-19. In: Christian H, Recker J, WK, editors. *Digital responses to COVID-19: digital innovation, transformation, and entrepreneurship during pandemic outbreaks.* Springer International Publishing; 2021. pp. 73–85. [https://doi.org/10.1007/978-3-030-66611-8\\_6](https://doi.org/10.1007/978-3-030-66611-8_6).
12. Naslund JA, Bondre A, Torous J, Aschbrenner KA. Social media and mental health: benefits, risks, and opportunities for research and practice. *J Technol Behav Sci.* 2020;5:245–57. <https://doi.org/10.1007/s41347-020-00134-x>.
13. Shaughnessy K, et al. Using geolocated social media for ecological momentary assessments of emotion: innovative opportunities in psychology science and practice. *Can Psychol.* 2018;59:47–53. <https://doi.org/10.1037/cap0000099>.
14. Gupta A, Katarya R. Social media based surveillance systems for healthcare using machine learning: a systematic review. *J Biomed Inform.* 2020. <https://doi.org/10.1016/j.jbi.2020.103500>.
15. Şerban O, Thapen N, Maginnis B, Hankin C, Foot V. Real-time processing of social media with SENTINEL: a syndromic surveillance system incorporating deep learning for health classification. *Inf Process Manage.* 2019;56:1166–84. <https://doi.org/10.1016/j.ipm.2018.04.011>.
16. Gruebner O, et al. Big data opportunities for social behavioral and mental health research. *Soc Sci Med.* 2017;189:158–66. <https://doi.org/10.1016/j.socscimed.2017.07.018>.
17. Gruebner O, et al. A novel surveillance approach for disaster mental health. *PLoS ONE.* 2017. <https://doi.org/10.1371/journal.pone.0181233>.
18. Chen X, Sykora MD, Jackson TW, Elayan S, Munir F. Tweeting your mental health: exploration of different classifiers and features with emotional signals in identifying mental health conditions. In *HICCS Hawaii International Conference on Computer Systems.* 2018.
19. Sykora MD, Jackson TW, O'Brien A, Elayan S. National security and social media monitoring: a presentation of the emotive and related systems. In *Proceedings—2013 European Intelligence and Security Informatics Conference, EISIC 2013*, pp. 172–175. <https://doi.org/10.1109/EISIC.2013.38>.
20. Elayan S, et al. The stresscapes ontology system: detecting and measuring stress on social media. In *ECSM-2020—7th European Conference on Social Media*, Larnaca, Cyprus. 2020.
21. Gruebner O, et al. Mental health surveillance after the terrorist attacks in Paris. *Lancet.* 2016;387:2196–7. [https://doi.org/10.1016/S0140-6736\(16\)30602-X](https://doi.org/10.1016/S0140-6736(16)30602-X).
22. Edry T, et al. Real-time geospatial surveillance of localized emotional stress responses to COVID-19: a proof of concept analysis. *Health Place.* 2021. <https://doi.org/10.1016/j.healthplace.2021.102598>.
23. Saha K, Torous J, Caine ED, de Choudhury M. Psychosocial effects of the COVID-19 pandemic: large-scale quasi-experimental study on social media. *J Med Internet Res.* 2020;22(11):e22600. <https://doi.org/10.2196/22600>.
24. Saha K, et al. A social media study on the effects of psychiatric medication use. In *Proceedings of the Thirteenth International AAAI Conference on Web and Social Media.* 2019.
25. Guntuku SC, et al. Tracking mental health and symptom mentions on Twitter during COVID-19. *J Gen Intern Med.* 2020;35:2798–800. <https://doi.org/10.1007/s11606-020-05988-8>.
26. Valdez D, Ten Thij M, Bathina K, Rutter LA, Bollen J. Social media insights into US mental health during the COVID-19 pandemic: longitudinal analysis of twitter data. *J Med Internet Res.* 2020;22(12):e21418. <https://doi.org/10.2196/21418>.
27. Li D, Chaudhary H, Zhang Z. Modeling spatiotemporal pattern of depressive symptoms caused by COVID-19 using social media data mining. *Int J Environ Res Public Health.* 2020;17:1–23. <https://doi.org/10.3390/ijerph17144988>.
28. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Depression Prim Care.* 2002. <https://doi.org/10.3928/0048-5713-20020901-06>.
29. Zhang B, Zaman A, Silenzio V, Kautz H, Hoque E. The relationships of deteriorating depression and anxiety with longitudinal behavioral changes in google and youtube use during COVID-19: observational study. *JMIR Ment Health.* 2020;7(11):e240127. <https://doi.org/10.2196/24012>.
30. Choi D, et al. Development of a machine learning model using multiple, heterogeneous data sources to estimate weekly US suicide fatalities. *JAMA Netw Open.* 2020. <https://doi.org/10.1001/jamanetworkopen.2020.30932>.

31. Back MD, et al. Facebook profiles reflect actual personality, not self-idealization. *Psychol Sci.* 2010;21:372–4. <https://doi.org/10.1177/0956797609360756>.
32. BBC News. Facebook Files: 5 things leaked documents reveal. <https://www.bbc.com/news/technology-58678332>. 2021.
33. Pagoto S, Nebeker C. How scientists can take the lead in establishing ethical practices for social media research. *J Am Med Inform Assoc.* 2019;26:311–3. <https://doi.org/10.1093/jamia/ocy174>.
34. McKee R. Ethical issues in using social media for health and health care research. *Health Policy.* 2013;110:298–301. <https://doi.org/10.1016/j.healthpol.2013.02.006>.
35. Benton A, Coppersmith G, Dredze M. Ethical research protocols for social media health research. In *Proceedings of the First Workshop on Ethics in Natural Language Processing.* 2017, pp 94–102.
36. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. *Curr Opin Psychol.* 2016;9:77–82. <https://doi.org/10.1016/j.copsyc.2016.01.004>.
37. Zimmer M. Addressing conceptual gaps in big data research ethics: an application of contextual integrity. *Soc Med Soc.* 2018. <https://doi.org/10.1177/2056305118768300>.
38. Martinez-Martin N, et al. Ethics of digital mental health during COVID-19: crisis and opportunities. *JMIR Ment Health.* 2020;7(12):e23776. <https://doi.org/10.2196/23776>.

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