The use of social media in public health surveillance

Isaac Chun-Hai Fung,[a] Zion Tsz Ho Tse[b] and King-Wa Fu[c]

Correspondence to Isaac Chun-Hai Fung (email: cfung@georgiasouthern.edu).

Social media are broadly understood as a set of online activities that facilitate interpersonal communication, information sharing, collaboration or crowdsourcing among online users. They have become a global phenomenon with over two thirds of worldwide adult Internet users being active on social networking sites in 2014.[1]

Social media are increasingly harnessed for public health and can be used as communication tools to disseminate disease risks and interventions and to promote healthy lifestyles and health policies. There is also the potential use of social media as data sources for public health surveillance. While social media will likely never replace traditional data sources for disease surveillance, they can provide complementary information. However, social media data are, in essence, observational data of online communications and were not designed for public health purposes. Analyses of social media data are subject to limitations that are generally associated with observational studies, i.e. possible confounding factors and no causal conclusion.

Following is a brief overview of some of the uses of social media data for public health surveillance and some of the data's strengths and limitations.

USING SOCIAL MEDIA FOR PUBLIC HEALTH SURVEILLANCE

There are three major applications for social media in public health surveillance: epidemiologic monitoring and surveillance, situational awareness during emergency response and communication surveillance (**Table 1**).

Epidemiologic monitoring and surveillance

For epidemiologic monitoring and surveillance, social media can be used to perform three specific functions:

Monitoring and retrieving official information

Public health officials use social media to monitor official information released by foreign authorities and to monitor domestic official accounts as these can be more timely, which is important in emergency responses (**Table 1**).[2,3]

Disease detection

Social media and other population-based digital platforms provide additional data sources for public health surveillance to detect disease outbreaks and estimate disease incidence. Syndromic surveillance can be undertaken by detecting symptoms disclosed by individuals on social media for non-public health purposes by either human readers or computer algorithm,[4] or through participatory epidemiology where applications allow participants to self-report their symptoms to disease-specific digital surveillance systems.[5,6] Diseases may also be detected via event-based surveillance as unofficial information or rumours about a new disease may circulate on social media, as in the case of a medical record of an H7N9 patient uploaded to Weibo in 2013.[7] Print media, radio and television may pick these up and generate news stories that are also circulated on social media and may be detected by event-based surveillance systems that monitor digital news feeds for disease news (e.g. HealthMap).[8] Digital data sources also provide epidemiologists with additional means to detect, investigate and verify outbreaks.

a   Department of Epidemiology, Jiann-Ping Hsu College of Public Health, Georgia Southern University, Statesboro, Georgia 30460-8015, United States of America.
b   College of Engineering, The University of Georgia, Athens, Georgia 30602, United States of America.
c   Journalism and Media Studies Centre, The University of Hong Kong, Hong Kong Special Administrative Region, China.

Table 1. **Summary of the uses of social media in public health surveillance**

Application	Purpose and scenario	Public health information to retrieve, detect or predict	Data targeted	Function of social media	Examples
1. Epidemiologic monitoring and surveillance					
(a) Monitoring official information	To monitor official information	Disease incidence and other case details	Links to original sources of official data	News feed	Retrieval of official information via Weibo during the 2013 H7N9 outbreak[2,3]
(b) Disease detection – syndromic surveillance	To detect outbreaks and to estimate disease incidence	Disease incidence and other case details	Self-reported symptoms	Syndromic surveillance	Twitter tweets of self-disclosed symptoms of influenza infection;[4] self-reports of symptoms via specialized apps[5,6]
(c) Disease detection – event-based surveillance	To detect outbreaks and to estimate disease incidence	Disease incidence and other case details	Media reports, unofficial information or rumours used as proxy measurable outcomes	Event-based surveillance (epidemic intelligence)	Unofficial information released on Weibo about an H7N9 patient;[7] systems that pick up news related to health events (e.g. HealthMap)[8]
(d) Timely estimates and forecasting of disease incidence	To provide timely estimates of current disease incidence or forecast future disease incidence	Disease incidence: start, peak and intensity	Social media text with keywords (diseases or symptoms) that correlate with disease incidence	As data sources for timely estimates or forecasts of disease incidence	US seasonal influenza estimates using Twitter data;[4] disease estimation and forecasts using Wikipedia access log;[9] US seasonal influenza forecast using Google Flu Trends[11]
2. Situational awareness					
(a) Surveillance for situational awareness	Humanitarian crises, usually natural disasters, e.g. typhoons and earthquakes	Reported needs (e.g. water supply and shelter)	Self-reported humanitarian needs	Information feed on humanitarian needs	Earthquake and tsunami in East Japan[10] and earthquake in Haiti[11]
3. Communication surveillance					
(a) Global awareness	To measure social media users' reactions to an outbreak situation	Media news reports, rumours, sentiments, awareness	User-generated data that reflect their knowledge, attitudes and perception of public health events	Monitoring of the general public's awareness and perception	Awareness of Ebola;[14] sentiment towards influenza vaccine[15]
(b) Specific reactions	To measure social media users' reactions to health promotion messages or events	Reception of particular public health messages	User generated data that are reactions to particular public health messages	Monitoring of the general public's reaction to specific public health messages	Breast cancer awareness month[16]

Timely estimates and forecasting of disease incidence

Epidemiologists are exploring ways to use social media and other digital data to provide timely estimates and forecasts of disease incidence. For example, Twitter data pertaining to influenza could facilitate timely incidence estimates as they were found to correlate with seasonal influenza data in the United States of America (USA).[4] Wikipedia access log data were also found to have potential for forecasting certain infectious diseases in some countries.[9] However, Google Flu Trends underperformed in comparison with the USA sentinel influenza-like illness surveillance system (ILI-Net) and a New York City syndromic surveillance system.[10] Advanced forecasting

methods are also under development, with some using digital data as experimental inputs.[11]

Situational awareness during emergency response

Social media can be used following natural or man-made disasters to increase situational awareness of humanitarian crises. Individuals in distress can use social media to seek help and to connect with family, friends and emergency responders. The authorities can use social media to identify individuals in distress and to respond accordingly. Nongovernmental organizations can also use social media to track and map the needs of

displaced people, as seen with the 2011 earthquake and tsunami in Japan,[12] and the 2010 Haitian earthquake.[13]

Communication surveillance

Global awareness

Social media data can also provide measures of global awareness of disease outbreaks. Complementary to more traditional methods, social media trends can help to quantify changes in disease awareness,[14] and sentiments towards treatments and preventive interventions.[15]

Reaction to public health campaigns and messages

Analyses of social media data pertaining to specific health-promotion events can provide useful insights to public health professionals as they evaluate their campaigns.[16]

STRENGTHS AND LIMITATIONS OF SOCIAL MEDIA DATA

In societies where penetration of social media is high, analyses of social media data can be compared to large-scale observational population-based epidemiologic studies. Public health researchers can have access to user-generated content from millions of users worldwide. However, given the inherent observational nature of social media data, their analyses are subject to many limitations. For example, selection bias may exist as social media users and non-users may differ. Privacy settings on social media may restrict data access to some individuals. Personal information of social media users is often self-disclosed and difficult to verify. There could also be observer effects, as the awareness of Internet surveillance and the fear of retribution may render social media users unwilling to share epidemiologic information, especially in countries where real-name registration of social media is required.[17]

The balance between sharing public health information and protecting citizens' privacy remains an ethical challenge for public health agencies. Given these concerns, many public health-related social media studies have been conducted using only publicly accessible data. However, their generalizability remains a question because often people prefer to communicate health-related issues privately.

CONCLUSIONS

Social media offer both opportunities and challenges to public health professionals. Social media data can allow public health officials to monitor information, detect potential outbreaks, forecast disease trends, monitor emergency situations and gauge disease awareness and reactions to official health communications. Questions remain as to how to best analyse social media data for public health surveillance. Public health agencies need to clearly define the purposes of the surveillance systems, the scope of social media data to be used and how the data should be analysed.

References

1. Mander J. *GWI Social: GlobalWebIndex's quarterly report on the latest trends in social networking (Q4 2014)*. London, GlobalWebIndex, 2015.

2. Fung IC-H, Wong KK. Efficient use of social media during the avian influenza A(H7N9) emergency response. *Western Pacific Surveillance and Response Journal*, 2013, 4(4):1–3. doi:10.5365/wpsar.2013.4.3.005 pmid:24478916

3. Zhang EX et al. Leveraging social networking sites for disease surveillance and public sensing: the case of the 2013 avian influenza A(H7N9) outbreak in China. *Western Pacific Surveillance and Response Journal*, 2015, 6(2). doi:10.5365/wpsar.2015.6.1.013

4. Broniatowski DA, Paul MJ, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012–2013 influenza epidemic. *PLoS ONE*, 2013, 8:e83672. doi:10.1371/journal.pone.0083672 pmid:24349542

5. Freifeld CC et al. Participatory epidemiology: use of mobile phones for community-based health reporting. *PLoS Medicine*, 2010, 7:e1000376. doi:10.1371/journal.pmed.1000376 pmid:21151888

6. Chunara R et al. Estimating influenza attack rates in the United States using a participatory cohort. *Scientific Reports*, 2015, 5:9540. doi:10.1038/srep09540 pmid:25835538

7. Salathé M et al. Influenza A (H7N9) and the importance of digital epidemiology. *The New England Journal of Medicine*, 2013, 369:401–404. doi:10.1056/NEJMp1307752 pmid:23822655

8. Brownstein JS, Freifeld CC. HealthMap: the development of automated real-time internet surveillance for epidemic intelligence. *Euro Surveillance: European Communicable Disease Bulletin*, 2007, 12(11):E071129.5. pmid:18053570

9. Generous N et al. Global disease monitoring and forecasting with Wikipedia. *PLoS Computational Biology*, 2014, 10:e1003892. doi:10.1371/journal.pcbi.1003892 pmid:25392913

10. Olson DR et al. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. *PLoS Computational Biology*, 2013, 9:e1003256. doi:10.1371/journal.pcbi.1003256 pmid:24146603

11. Shaman J et al. Real-time influenza forecasts during the 2012–2013 season. *Nature Communications*, 2013, 4:2387. doi:10.1038/ncomms3837 pmid:23982432

12. Peary BDM, Shaw R, Takeuchi Y. Utilization of social media in the east Japan earthquake and tsunami and its effectiveness. *Journal of Natural Disaster Science*, 2012, 34:3–18. doi:10.2328/jnds.34.3

13. Zook M et al. Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Medical & Health Policy*, 2010, 2:7–33. doi:10.2202/1948-4682.1069

14. Fung IC-H et al. Ebola and the social media. *Lancet*, 2014, 384:2207. doi:10.1016/S0140-6736(14)62418-1 pmid:25625391

15. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS Computational Biology*, 2011, 7:e1002199. doi:10.1371/journal.pcbi.1002199 pmid:22022249

16. Thackeray R et al. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. *BMC Cancer*, 2013, 13:508. doi:10.1186/1471-2407-13-508 pmid:24168075

17. Fu K-W, Chan CH, Chau M. Assessing censorship on microblogs in China: discriminatory keyword analysis and the real-name registration policy. *Internet Computing, IEEE*, 2013, 17:42–50. doi:10.1109/MIC.2013.28