

Routes of communication: understanding the information flows that shape public and official reaction to COVID-19

1 Project overview

Information related to health risks and healthy behaviour is typically generated by scientists and distilled into recommendations by public-health agencies. From there it is often transmitted by mass media to the public. Misinformation, generated by careless or irresponsible scientists, pseudo-scientists, or product-marketers can follow a similar route. In the age of the internet, these routes remain important, but the public also has easy direct access to information from public-health agencies, and to many scientific papers, including unvetted preprints. The public also has an expanded ability to interpret, transmit, and amplify messages through social media. More recently still, scientists and agencies have become active on social media as well.

During disease outbreaks of international concern, like 2003 SARS, 2009 H1N1, 2014 Ebola, or the current COVID-19 outbreak, this process is both compressed and amplified. The stakes also become higher, because public behaviour directly affects the spread of infectious disease: people who avoid large gatherings may slow disease spread, while people who flee infected areas may accelerate it, for example. Excessive fear of disease spread can have severe economic effects, and may also lead to bias and discrimination against groups seen as linked to the disease, or identified as “others”.

The proposed research will study how information flows between forums, including scientific publications; governmental policies and agency recommendations; and mass and social media – and investigate how it affects public perceptions and behaviours. Our interdisciplinary group will analyze these flows by combining textual and contextual analysis; AI-assisted human-supervised data mining; time-series analysis; and dynamical modeling. We will study how good information competes with misinformation, and look for factors correlated with successful spread of good information. We will gather information on communication and surrogates for behaviour from a wide range of sources. Sources about communication will include agency websites; preprint servers and publicly available scientific journals; major newspaper websites; and social-media platforms. Surrogates for public perceptions and behaviour will include data from open social-media platforms twitter and Weibo; google trends; publicly available box-office information for movies and major sporting leagues; and publicly available travel information; and information about cancellations and shortages (for example of face masks or pharmaceutical or pseudo-pharmaceutical products) from our textual analysis.

The project will be organized around three Research Questions (RQs), all based on the Social and policy countermeasures research area from the funding opportunity:

RQ1 How does information (and misinformation) travel between scientists, public-health workers, mass media and social media? (*based on funding opportunity research sub-area: “cultural dimensions of the epidemic”*)

RQ2 How does communication affect public behaviour and the course of the outbreak? (*research sub-area: “feasibility and effectiveness of public health response”*)

RQ3 How can scientists and policy-makers evaluate and improve the effectiveness of their communication? (*research sub-area*: “strategies to combat misinformation”)

2 Background

Public-health communication is a balancing act. Officials are often caught between the need to be heard, and the danger of causing panic. This problem is particularly acute in the case of an infectious disease outbreak, since the presence of a novel pathogen increases both the importance of being heard and the danger that the public will over-react.

In the case of COVID-19, scientists are still scrambling to understand the pathogen’s biology; public-health workers are scrambling to decide on the best recommendations and policy decisions given current knowledge at any given time; and the mass media is scrambling to understand the situation and decide how best to communicate with the public.

There are other complicating factors. An outbreak of global concern represents an opportunity for mainstream and peripheral media, and for social-media actors to increase their “clicks” and “likes” and therefore prestige and/or profitability. These motivations work against the balancing act, and instead favor over-simplification and sensationalization.

It is known that traditional media can strongly influence public perceptions, creating fear by overestimating risk during the SARS outbreak [1] and the influenza pandemic [2], or feeding into bias – e.g., anti-Chinese bias during both the SARS crisis [?] and the current outbreak. Media is also the tool public health authorities rely on to promote their concerns and recommendations during outbreaks. Understanding media effects on disease spread (e.g., media attention increases self-protection) can help enhance epidemic forecasting and preventive measures to slow the disease spread [3].

While traditional news media (including online presence) remains influential, social media plays an increasingly important role in shaping how we communicate and understand information [4]. Social media can play a positive role spreading good information, [5–7], but may also spread misinformation and feed bias [?, 8]. Since the initial reports, a cluster of acute severe respiratory disease (COVID-19) and the potential for global spread, there has been widespread discussion and dissemination of information through social media [?, 9, 10].

3 Methods and feasibility

3.1 Data

Science We will develop systematic search and screening strategies to extract relevant peer-reviewed publications from Google Scholar and PubMed. To account for the strong influence of preprints early in the epidemic [?], we will also include preliminary scientific findings posted on medRxiv through 31 March 2020. We will index these papers and track their appearances in mass media and social media; we will also track which of the preprints are published after peer review.

Public health recommendations We will collect and analyze reports, guidelines and recommendations available from the World Health Organization and from the central disease control agency of each of our focal countries. It is worth noting that all of these agencies have launched special COVID-19 web pages.

Mass media We will use the NexisUni search engine (via McMaster University) and OriProbe Information Services to collect articles relevant to the outbreak, going back to the outbreak start in December 2019, and continuing throughout the grant period. We will focus on the top English- and Mandarin-language newspapers (taking both circulation and online access into account) from Canada, China, England, Singapore, Taiwan, and the USA. We will include the top Mandarin-language newspapers in both Canada and the USA.

Social media We will efficiently collect data from Twitter and Weibo by purchasing API access, using data going back to November 2019 – before the epidemic started – to give a baseline for comparison.

Public response Twitter (and Weibo) data will give us information not only on information flows, but also on public interest, attitudes and topics being discussed on the social media. We will also probe public interest and concern using publicly available data from GoogleTrends, which tabulates frequency of searches (by search times and topics) in various regions across the world [11,12]. Economic data relating to the outcome of the COVID-19 outbreaks will be gathered as reference for contextual analysis (see below): for example, travel, movies box offices, cancellation of public events.

3.2 Analysis

Textual analysis To investigate how information/misinformation travels and how communication affects public responses (e.g., attitudes), we will use state-of-the-art machine learning and natural language processing (NLP) techniques. We have two directions. First, to investigate how information travels, we plan to develop codebooks to manually annotate random samples of articles/messages from scientific papers, government recommendations, mass media and social media. Codebooks will contain both themes and frames relevant to our analysis. For example, correct or incorrect information (misinformation or not?) will be annotated. Then, using the annotated data as training data, supervised machine learning algorithms such as logistic regression, Support Vector Machine, and neural network will learn patterns of textual and content features to predict the label(code?). Second, to study public responses, we will leverage NLP techniques such as aspect-based sentiment analysis and topic modeling. For example, ABSApp, a state-of-the-art aspect-based sentiment extraction tool, allows extracting important aspects of the target(?) situation and detecting sentiment towards the aspects, with relatively less human involvement. Latent Dirichlet Allocation (LDA) and its variants, popularly used topic modeling techniques, help identify topics in large texts. These approaches will allow us to perform a large scale study (using large data). For all the textual analysis processes, computer scientists will iterate collaboratively with coders and subject-matter experts to build codebooks that are consistent with study aims, and to interpret findings from applying AI algorithms. The findings would be shared with public-health practitioners to assist with counter-messaging strategies.

Time-series analysis We will use cross-correlation analyses to look for indicators that information is moving from one communication forum to another; that events (like disease spread or public behavior) are affecting communication; or that communication is affecting

events. Cross-correlation analysis is complicated and prone to false-positive results. Importantly, therefore, we will be able to use the cross-correlation analysis to generate hypotheses that can be checked by more detailed textual analysis. For example, if we hypothesize that tweets about fatalities are being driven at a certain time and place by mass media, or by government policies, we can sample from those tweets and examine them for detailed information or citations; if we hypothesize that a trend in self-isolation is driven by social media, we can search for mass media stories that interview people about their motivation. The ability to compare large-scale trends with detailed texts should amplify our pattern to detect and confirm patterns.

Dynamical modeling Dynamical modeling provides the link between individual events and emergent phenomena. We will make a range of simple dynamical models to further probe our time-series results by asking what mechanisms may underlie our observed connections, and what these connections might imply for the future. Dynamical models will allow us to explore hypotheses about what factors affect behaviour, and also to explore new hypotheses about how changes in behaviour are likely to loop back to disease transmission or to panic responses that might lead to shortages or to impacts on regional economies or the global economy. ((naveed: This is not clear to me what we are trying to model here and how this will link back with the other themes.))

Synthesis Effective health communication, including dissemination of good information and countering misinformation, is key to outbreak management [13] and consistent recommendation [13, 14]. We will use techniques from content analysis to combine results from our analysis techniques above to formulate hypotheses about what factors lead to effective communication. In particular, we will identify cases where good information did or did not out-compete bad information. When information from public-health agencies spreads effectively we will also evaluate our behavioural proxies to ask when it led to a calibrated reaction from the public (as opposed to over- or under-reaction). We will make use of previous studies to evaluate [?, ?, ?, ?] and develop [?, 13, 15] strategies for effective health communication.

3.3 deliverables

- software ((naveed: Describe what kind of software. I was envisioning a tool that runs on Twitter stream data and classifies tweets based on contents as well as significance in terms of influence (how much it has potential to spread based on following network or influencer. See paper for classification of influencers. <https://bmcpublihealth.biomedcentral.com/articles/10.1186/s12889-019-6747-8>))
- Communication: Writing papers is certainly one, but we should be tweeting and etc. What are communication platforms we should be communicating in? ((cs: attending conference, meeting with policy makers with our findings; possible WHO network?)) ((naveed: It depends on objective, would we be doing counter messaging or public health. The way I was envisioning, creating a tool to identify and alert public health communication team and they addressing it through tweets or other means.)) ((de: Perhaps combine the tweets with a

blog, which would engage different audiences. The tweets could summarize more elaborate blog posts. I have no experience writing either a blog or tweets myself...))

4 Research Setting & Personnel

The principal applicants have a long history of influential research on disease outbreaks:

The research will principally take place at McMaster University. Nominated principal applicant Dr. David J.D. **Earn** (6 hours per week) led the creation of the International Infectious Disease Data Archive and has expertise in gathering and curating infectious disease data, and in dynamical modelling, including modelling the influence of individual decision-making on epidemic dynamics. Principal applicant Dr. Jonathan **Dushoff** (4 hours per week) is an internationally recognized expert in infectious disease modelling, has extensive experience with statistical frameworks for fitting models to data, and has been involved in the Ebola challenge and other forecasting projects. Co-Applicant Dr. Benjamin **Bolker** (5 hours per week) is a highly accomplished ecological statistician with extensive experience in spatial-dynamical modeling, statistical modeling and statistical software. Co-Applicant Dr. Naveed Z. **Janjua** (5 hrs week) leads data and analytic services at the BCCDC and was involved in the 2009 H1N1 pandemic response; during pandemic, lead or contributed to studies on immuno-epidemiology of pandemic H1N1, household transmission, modelling, effect of prior seasonal vaccine receipt on pandemic H1N1 infection risk and pandemic vaccine effectiveness. Co-Applicant Dr. Chyun-Fung **Shi** (40 hours per week), post-doctoral researcher, has ... Co-Applicant Dr. Michael **Li** (5 hours per week), post-doctoral researcher, has focused his research on epidemic forecasting and is experienced working with large databases.

5 Research Time line

((mli: not sure if we need this section, but CIHR want to see we are working on things and are ready to do more work.)) ((naveed: I think we need to see what we could start quickly. HyeJu our postdoc has already started working on getting twitter data and starting to analyze it to make sense. We should say that are ready to get started.))

5.1 Year 1

5.2 Year 2

5.3 Potential Outcomes

Transfer findings to peer-reviewed publications. In addition to a grand one paper, we plan to transfer findings into subgroups, such as by affected country (North America, Asia-Taiwan, H.K. and Singapore), types of media (news media and social media (twitter)).

- Potential to contribute to the global response to COVID-19
- Social and policy countermeasures and Global Coordination Mechanism

We could say; We will work with public health teams to inform their communicate strategies tackling misinformation and design a tool for real-time monitoring and tackling misinformation.

We will also provide data on magnitude of discriminatory information and its geographical origin to tackle discrimination.))((naveed: Data for public health response. Our provincial health officer is asking for data and today there was a visit from Federal Health Minister and BC Health Minister to Richmond related to affect on Chinese community.

We could say; We will work with public health teams to inform their communicate strategies tackling misinformation and design a tool for real-time monitoring and tackling misinformation.

We will also provide data on magnitude of discriminatory information and its geographical origin to tackle discrimination.))

6 Challenges and Mitigation Strategies

Social media in China is to included due to data availability. Social media such as twitter and Facebook are barred in China and WeChat does not share their database. Yet, there are information and coverage about the outbreak inside China in the media. We will categorize people's reaction and understanding of COVID-19 on the news content or the tweets to supplement our results.

((naveed: What about Weiboo. Good to highlight that data may not be available or limited in scope. Other mechanism of collecting that data e.g., web crawler for data available in public domain.))

Data curation We will go back and document clearly all the policy changes and case definitions.

- data collected from National Health Commission
- Figure out how to use the data effectively (e.g. we are not using death, severity categorizations, number of tested, number of positive, and etc)
- Case definitions
- Media content
- Social media platforms
- Google trends

Analysis/ Pipelineing/ mainstreaming Info delays, misinformation and miscommunications

Communication how people are interpreting

233 7 Conclusion/summary

References

- [1] T. R. Berry, J. Wharf-Higgins, and P. J. Naylor. SARS wars: an examination of the quantity and construction of health information in the news media. *Health Commun*, 21:35–44, 2007.
- [2] J. M. Tchenche, N. Dube, C. P. Bhunu, R. J. Smith, and C. T. Bauch. The impact of media coverage on the transmission dynamics of human influenza. *BMC Public Health*, 11 Suppl 1:S5, 2011 Feb 25.
- [3] L. Kim, S. M. Fast, and N. Markuzon. Incorporating media data into a model of infectious disease transmission. *PLoS One*, 14:e0197646, 2019.
- [4] J. Liu, L. Siegel, L. A. Gibson, Y. Kim, S. Binns, S. Emery, and R. C. Hornik. Toward an Aggregate, Implicit, and Dynamic Model of Norm Formation: Capturing Large-Scale Media Representations of Dynamic Descriptive Norms Through Automated and Crowdsourced Content Analysis. *J Commun*, 69:563–588, 2019 Dec.
- [5] C. H. Basch and G. C. Hillyer. Skin cancer on Instagram: implications for adolescents and young adults. *Int J Adolesc Med Health*, 2020 Feb 7.
- [6] M. Sun, L. Yang, W. Chen, H. Luo, K. Zheng, Y. Zhang, T. Lian, Y. Yang, and J. Ni. Current status of official WeChat accounts for public health education. *J Public Health (Oxf)*, 2020 Jan 23.
- [7] N. Ahmed, S. C. Quinn, G. R. Hancock, V. S. Freimuth, and A. Jamison. Social media use and influenza vaccine uptake among White and African American adults. *Vaccine*, 36:7556–7561, 2018 Nov 26.
- [8] W. Y. S. Chou, A. Oh, and W. M. P. Klein. Addressing Health-Related Misinformation on Social Media. *JAMA*, 320:2417–2418, 2018 Dec 18.
- [9] C. Chew and G. Eysenbach. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One*, 5:e14118, 2010 Nov 29.
- [10] L. Tang, B. Bie, S. E. Park, and D. Zhi. Social media and outbreaks of emerging infectious diseases: A systematic review of literature. *Am J Infect Control*, 46:962–972, 2018 Sep.
- [11] J. Bousquet, I. Agache, J. M. Anto, K. C. Bergmann, C. Bachert, I. Annesi-Maesano, P. J. Bousquet, G. D’Amato, P. Demoly, G. De Vries, E. Eller, W. J. Fokkens, J. Fonseca, T. Haahtela, P. W. Hellings, J. Just, T. Keil, L. Klimek, P. Kuna, K. C. Lodrup Carlsen, R. Mosges, R. Murray, K. Nekam, G. Onorato, N. G. Papadopoulos, B. Samolinski, P. Schmid-Grendelmeier, M. Thibaudon, P. Tomazic, M. Triggiani, A. Valiulis, E. Valovirta, M. Van Eerd, M. Wickman, T. Zuberbier, and A. Sheikh. Google Trends terms reporting rhinitis and related topics differ in European countries. *Allergy*, 72:1261–1266, 2017 Aug.

-
- [12] N. Mahroum, N. L. Bragazzi, F. Brigo, R. Waknin, K. Sharif, H. Mahagna, H. Amital, and A. Watad. Capturing public interest toward new tools for controlling human immunodeficiency virus (HIV) infection exploiting data from Google Trends. *Health Informatics J*, 25:1383–1397, 2019 Dec.
- [13] J. Song, T. M. Song, D. C. Seo, D. L. Jin, and J. S. Kim. Social Big Data Analysis of Information Spread and Perceived Infection Risk During the 2015 Middle East Respiratory Syndrome Outbreak in South Korea. *Cyberpsychol Behav Soc Netw*, 20:22–29, 2017 Jan.
- [14] A. W. Ohlrogge and L. S. Suggs. Flu vaccination communication in Europe: What does the government communicate and how? *Vaccine*, 36:6512–6519, 2018 Oct 22.
- [15] M. A. Bekalu, C. A. Bigman, R. F. McCloud, L. K. Lin, and K. Viswanath. The relative persuasiveness of narrative versus non-narrative health messages in public health emergency communication: Evidence from a field experiment. *Prev Med*, 111:284–290, 2018 Jun.