

# How infodemic intoxicates public health surveillance: from a big to a slow data culture

Arnaud Chiolero 

Population Health Laboratory,  
University of Fribourg, Fribourg,  
Switzerland

**Correspondence to**  
Professor Arnaud Chiolero,  
Population Health Laboratory  
(#PopHealthLab), University  
of Fribourg, Fribourg 1700,  
Switzerland;  
[achiolero@gmail.com](mailto:achiolero@gmail.com)

Received 29 November 2021  
Accepted 25 January 2022

## ABSTRACT

Too much data? Too much information? The COVID-19 pandemic has made the case. The WHO coined the term ‘infodemic’ to describe the issue of overabundance of information, including misinformation, disseminated in real time via multiple channels.<sup>1 2</sup> A related concept is ‘datademic’ to describe the overabundance of data. I argue in this essay that infodemic intoxicates public health surveillance and decision-making, and that we need to revisit how we conduct surveillance in the age of big data by fostering a slow data culture.

## WHY TOO MUCH DATA INTOXICATE PUBLIC HEALTH SURVEILLANCE

Surveillance is the ongoing systematic collection, analysis and interpretation of data, closely integrated with the timely dissemination of the resulting information to those responsible for preventing and controlling disease and injury.<sup>3</sup> Traditionally, it requires high-quality data which are collected for this purpose along well-defined methods. In the era of infodemic and big data, the access to different types of data has increased tremendously, offering new opportunities for surveillance. These new data, however, are not collected primarily for surveillance, often of relatively low (or not well-documented) quality, and, what is highly problematic for surveillance, of weak consistency across settings and time.<sup>3</sup> The consequences are the questionable quality and reliability of information derived from these data.

An additional problem is what is called the selectivity bias.<sup>4</sup> With some types of big data, it is difficult to define the source population from which they emerged; there is no well-defined source population. For instance, routinely collected data from healthcare providers are usually event based rather than population based, and the population using the services of these providers is changing, unpredictably, across time. These data are further exposed to surveillance bias and streetlight effect.<sup>3</sup> Defining the source population is also very difficult with data from social media networks or internet queries, and it is not clear of which population there are representative. As a result, how to infer information derived from these data to a target population is a serious challenge for decision-makers.

## MULTIPLICATION OF INFORMATION PRODUCERS

Beyond quality, volume is the other major problem that intoxicates surveillance activities. In the field

of public health surveillance, we are chronically complaining about the lack of data, for example, to estimate the burden of given diseases or to evaluate the effect of specific policies on population health. When available, these data have the advantage, however, of being collected to address relatively well-defined public health surveillance needs; they are designed for this purpose. In this time of infodemic, we are no short of data; on the contrary, we are actually overwhelmed by data and information which are on the one hand easily available but, on the other hand, not designed to address public health needs. New methods are needed to make them useful for decision-making.

Further, there is not only a multiplication of new data sources but also a multiplication of information producers (figure 1). Until recently, the production of surveillance information was tightly regulated, usually under the guidance of public health experts more or less trained in epidemiology and statistics, mandated by public agencies, and on which decision-makers could rely on. The health information production market is now open and deregulated, with data analysed by multiple people—with or without epidemiological expertise—who make their analyses available through different types of traditional and social media, and who compete with traditional information producers. The problem is that substantial resources are needed to handle this huge volume of information, decision-makers having to screen and sieve them to sort out what is useful for decision.

While poor quality and huge volume of data can lead to the production of information of questionable utility, a related problem is misinformation which can be defined as incorrect or false information that is shared with or without the intent to harm.<sup>5 6</sup> It is nurtured by several psychosocial mechanisms such as congeniality bias, motivated reasoning, doom-scrolling and negativity bias, and is amplified by the algorithms of social networks.<sup>7</sup> The spread of false information delays the adoption of effective public health measures, notably vaccination. Due to the scale of misinformation within the digital information ecosystem, a new—and resource-consuming—activity for public health agencies is tracking and countering misinformation.<sup>8 9</sup>

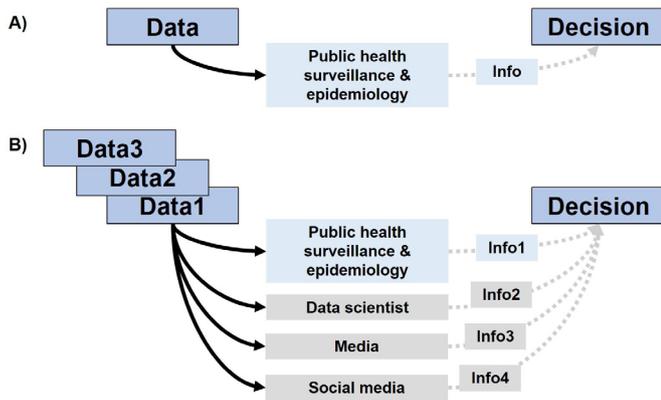
## FAILURE OF SCIENTISTS

Scientists have not helped substantially to mitigate the effects of infodemic and to prevent misinformation. Actually, they might have contributed to these trends. More than 500 000 scientific papers would



© Author(s) (or their employer(s)) 2022. No commercial re-use. See rights and permissions. Published by BMJ.

**To cite:** Chiolero A. *J Epidemiol Community Health* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jech-2021-216584



**Figure 1** Traditionally (A), public health surveillance is conducted using a given set of well-defined data and analysed by experts in surveillance and epidemiology to produce information designed, hopefully, to help decision. In the age of big data and infodemic (B), there is a not only a multiplication of data sources but also a multiplication of information producers. As a result, decision-makers have to account for a large volume of information produced by multiple players.

have been published on COVID-19 up to September 2021.<sup>10</sup> It has also become standard to practise ‘medicine by press release’ by providing information on a new treatment or a test directly through media. Hence, sensational and exaggerated rather than evidence-based and potentially useful research findings are spread.<sup>11</sup> Further, the mediatization of preprints has favoured the spread of not peer-reviewed information. Another trap for scientists is the academic militancy, blurring the boundary between science making and politics.<sup>12</sup>

More broadly, the COVID-19 pandemic magnifies how biomedical and public health science is evolving, with the mass production of low-value research and with much effort devoted to communication designed for academic self-promotion rather than for knowledge transfer.<sup>11</sup> In this era of infodemic, scientists should rather help decision-makers to deal with the mass of health information and sort out the good from the bad through, for example, misinformation-monitoring systems.<sup>13</sup> More than ever, there is a need for trustworthy gatekeepers of knowledge, working within an evidence-based framework, on which decision-makers can rely on.

### HOW TO COPE? MOVING FROM A BIG TO A SLOW DATA CULTURE

In the age of infodemic and big data, we need to slow the surveillance data and information production process. Making sense of these data and information requires time, as well as for making them useful for decision-makers. We call here for a ‘slow data’ culture defined as a shared way of collecting, analysing, and disseminating data and information in order to address explicitly the needs of public health decision-makers, and which can be fostered in three steps.

First is to admit that we need more than data for an efficient public health surveillance system. Indeed, we have to focus on the process to produce useful information matching the needs of decision-makers. Data whatever their scale are not enough; ‘big’ data do not speak by themselves more than ‘small’ data. In many surveillance systems, a large share of resources is used for data collection and analyses while more weight should be given to information dissemination and communication. We need to identify explicitly surveillance

needs, evaluate systematically what is the gain of considering extra data, and anticipate resources needed to handle these new data. We also need to strengthen the surveillance and containment of infodemic and misinformation, for example, by real-time tracking system of false COVID-19 claims and using trusted fact-checkers.<sup>13 14</sup>

Second, it is essential to re-enchant expertise and evidence. The pandemic has exposed strengths and weaknesses of scientists, and raised doubts on the value of reason and science.<sup>15</sup> Public health agencies, epidemiologists, and data scientists have however unique expertise and credibility to guide in the adequate use of data for surveillance.<sup>16</sup> More than ever, we need experts in public health surveillance with a clear understanding of the difference between evidence production and health policymaking,<sup>6</sup> at the interface between decision-making and data science, and working in trustworthy, science-grounded, accountable institutions. This is necessary to counter misinformation through the promotion and dissemination of credible information.<sup>14 17</sup>

Finally, it is necessary to strengthen skills of the society at large in the field of population health science. Epidemiology should be taught early in schools<sup>18</sup> and, within a consequentialist and evidence-based perspective,<sup>19</sup> epidemiologists have to be trained in knowledge translation and health communication.<sup>16</sup> In the age of big data and infodemic, increasing health data and epidemiological literacy at a population level is needed for citizen-centred and accountable public health surveillance systems.

**Twitter** Arnaud Chiolero @swissepi

**Contributors** AC wrote the manuscript and is responsible for the overall content as guarantor.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** This study does not involve human participants.

**Provenance and peer review** Commissioned; externally peer reviewed.

**Data availability statement** No data are available. Not applicable.

This article is made freely available for personal use in accordance with BMJ’s website terms and conditions for the duration of the covid-19 pandemic or until otherwise determined by BMJ. You may use, download and print the article for any lawful, non-commercial purpose (including text and data mining) provided that all copyright notices and trade marks are retained.

**ORCID iD**

Arnaud Chiolero <http://orcid.org/0000-0002-5544-8510>

**REFERENCES**

- 1 World Health Organization (WHO). Infodemic. Available: [www.who.int/health-topics/infodemic](http://www.who.int/health-topics/infodemic) [Accessed 28 Nov 2021].
- 2 Gabarron E, Oyeyemi SO, Wynn R. COVID-19-related misinformation on social media: a systematic review. *Bull World Health Organ* 2021;99:455–63.
- 3 Chiolero A, Buckeridge D. Glossary for public health surveillance in the age of data science. *J Epidemiol Community Health* 2020;74:612–616.
- 4 Beresewicz M, Lehtonen RT, Reis F. *An overview of methods for treating selectivity in big data sources*. Eurostat. Luxembourg: Publications Office of the European Union, 2018.
- 5 The Lancet Infectious Diseases. The COVID-19 infodemic. *Lancet Infect Dis* 2020;20:875.
- 6 Gallotti R, Valle F, Castaldo N, et al. Assessing the risks of ‘infodemics’ in response to COVID-19 epidemics. *Nat Hum Behav* 2020;4:1285–93.
- 7 Briand SC, Cinelli M, Nguyen T, et al. Infodemics: a new challenge for public health. *Cell* 2021;184:6010–4.
- 8 Gradońi KT, Holyst JA, Moy WR, et al. Countering misinformation: a multidisciplinary approach. *Big Data Soc* 2021;8:205395172110138.

- 9 Calleja N, AbdAllah A, Abad N, *et al*. A public health research agenda for managing Infodemics: methods and results of the first who Infodemiology conference. *JMIR Infodemiology* 2021;1:e30979.
- 10 Brainard J. No revolution: COVID-19 boosted open access, but preprints are only a fraction of pandemic papers. science. Available: [www.science.org/content/article/no-revolution-covid-19-boosted-open-access-preprints-are-only-fraction-pandemic-papers](http://www.science.org/content/article/no-revolution-covid-19-boosted-open-access-preprints-are-only-fraction-pandemic-papers) [Accessed 10 Jan 2022].
- 11 Ioannidis JPA. Coronavirus disease 2019: the harms of exaggerated information and non-evidence-based measures. *Eur J Clin Invest* 2020:e13223.
- 12 Chiolero A. Is science ever enough? Dare to play politics. *Lancet* 2021;397:23.
- 13 Scales D, Gorman J, Jamieson KH. The Covid-19 Infodemic - Applying the Epidemiologic Model to Counter Misinformation. *N Engl J Med* 2021;385:678–81.
- 14 Zarocostas J. How to fight an infodemic. *Lancet* 2020;395:676.
- 15 Pinker S. *Enlightenment now: the case for reason, science, Humanism, and progress*. New York: Viking Press, 2018.
- 16 Smith MY, Bahri P, Gaudino JA, *et al*. The role of epidemiologists in communicating SARS-CoV-2 evidence: a call for adopting standards. *Int J Epidemiol* 2021;50:1410–5.
- 17 Gruzd A, De Domenico M, Sacco PL, *et al*. Studying the COVID-19 infodemic at scale. *Big Data Soc* 2021;8:205395172110211.
- 18 Goldmann E, Stark JH, Kapadia F, *et al*. Teaching epidemiology at the undergraduate level: considerations and approaches. *Am J Epidemiol* 2018;187:1143–8.
- 19 Galea S. An argument for a consequentialist epidemiology. *Am J Epidemiol* 2013;178:1185–91.