Leveraging standardised data in response to the novel coronavirus outbreak

Derk Arts,[1] **Lauren McGill**[2]

[1] Castor EDC, Hoboken, NJ, USA
[2] Freelance Medical Writer, Ontario, Canada

Correspondence to:

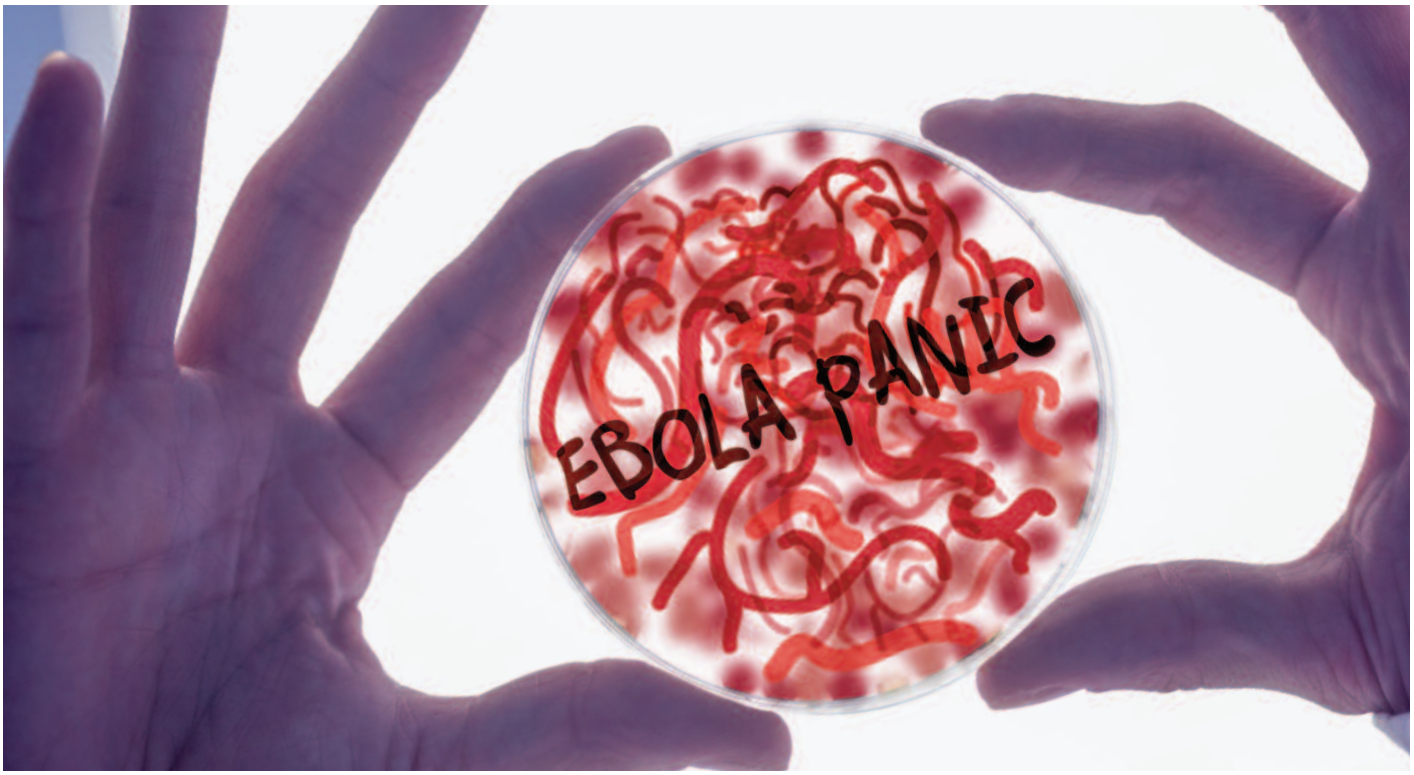
Derk Arts
221 River Street
9th floor
Hoboken, NJ 07030
USA
derk@castoredc.com

Abstract

As we are confronted by a new global health epidemic in the form of COVID-19, the challenges and opportunities of global data sharing come into sharp focus. Due to significant data collection and sharing issues during the 2013–2016 Ebola outbreak, the WHO recently called for improvements before the next public health emergency occurred. While it is too early to quantify the role of standardised data collection and sharing in containing the spread of COVID-19, it is possible to identify some of the data tactics used as part of the medical community's initial response.

Solid data are the best basis for public health action during an unfolding health emergency. Currently, the world is facing just such a crisis. On January 30, 2020, WHO declared the new contagious coronavirus COVID-19 a Public Health Emergency of International Concern. The virus spread rapidly from a single Chinese city through the entire country in just 30 days.[1] Over the following weeks, the world watched as an increasing number of countries reported confirmed cases, triggering government action and worldwide panic. While challenging, this latest global health crisis may prove key in the testing and implementation of new ways to collect, share, and aggregate data.

Castor vs. COVID-19



Lessons from the past

During an outbreak, data are our most valuable assets in the race to effective containment and finding a cure or vaccine. Effective governmental responses are only possible when there is accurate, real-time data available to base decisions on. At the beginning of the COVID-19 outbreak, WHO set clear expectations for better data collection and sharing than during previous outbreaks. In their statement on the second meeting of the "International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV)", WHO declared:

> *As this is a new coronavirus, and it has been previously shown that similar coronaviruses required substantial efforts to enable regular information sharing and research, the global community should continue to demonstrate solidarity and cooperation, in compliance with Article 44 of the IHR (2005), in supporting each other on the identification of the source of this new virus, its full potential for human-to-human transmission, preparedness for potential importation of cases, and research for developing necessary treatment.*[2]

Sadly, the need for "regular information sharing and research" was starkly highlighted during the disastrous handling of data during the 2013–2016 Western African Ebola virus epidemic. In hindsight, it's clear that there were many contributing factors to the difficulties encountered during that health emergency.[3]

Firstly, large pools of existing data from previous Ebola studies, a disease first discovered in 1976, had not been fully disseminated. Research had been conducted, but much of it was never published. When the recent severe outbreak triggered a rush to a cure, incomplete data led researchers down erroneous paths, wasting valuable time. During the multi-year outbreak itself, data were hap-haz-ardly collected and not standardised. There were large communication failures between affected countries and an unwillingness to share information with each other. Finally, data that were actually collected during the outbreak was not always standardised and was often shared inefficiently, with some researchers hesitating to share any data before they were ready for publication in an academic journal.

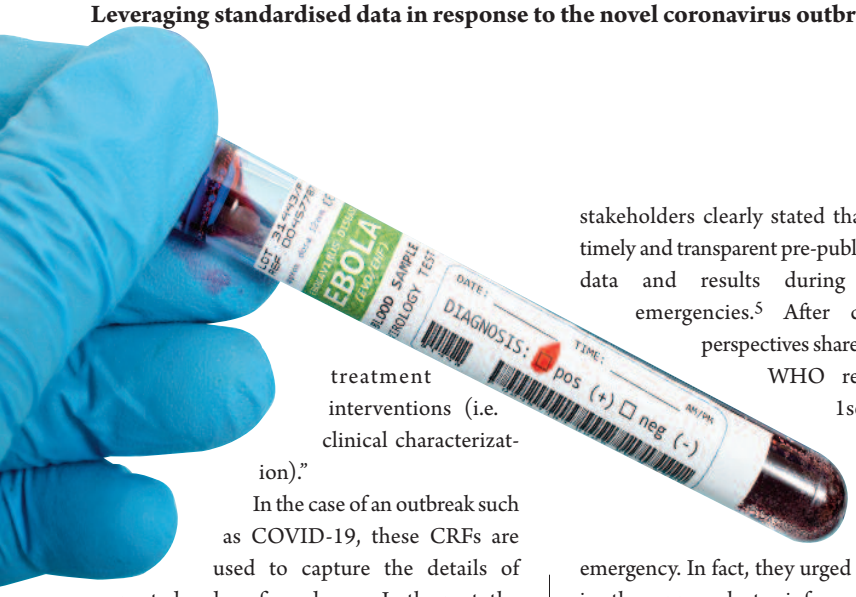
The full impact of the delays in gathering and sharing data during the Ebola outbreak may never be quantifiable. What we do know for certain is that we must rise to the challenge of effective data collection and sharing during current and future outbreaks.

The need for standardised data

In an epidemic, early data are key. It comes with a caveat though: data are only valuable if they are standardised. Amidst any epidemic, the goal is to assemble large amounts of accurate and usable data as quickly as possible. Swaths of data are used to help identify the causative agents; investigate and predict disease spread; define diagnostic criteria; and evaluate treatments and methods to contain further spread. Standardising data means using international recognised terminology for health concepts (e.g., SNOMED or LOINC) to annotate data, or in the very least, capture data in an agreed-upon data model so data from multiple research projects can be pooled and analysed in unison.

The full impact of the delays in gathering and sharing data during the Ebola outbreak may never be quantifiable.

In line with this, as the current COVID-19 outbreak began, WHO quickly provided technical guidance on how to conduct useful early investigations and provided the Global 2019-nCoV Clinical Characterization Case Record Form (CRF).[4] The CRF was "intended to provide member states with a standardized approach to collect clinical data in order to better understand the natural history of disease and describe clinical phenotypes and



treatment interventions (i.e. clinical characterization)."

In the case of an outbreak such as COVID-19, these CRFs are used to capture the details of suspected and confirmed cases. In the past, the reports were paper-based. More recently, medical researchers have used a combination of paper CRFs and electronic CRFs (eCRFs). But with a contagion as fast-moving as COVID-19, speed is key. Therefore, using eCRFs is the best way to hasten the aggregation of clinical data from around the world and accelerate the work of researchers.

In order to contribute to the global response, the medical data capture platform Castor (https://www.castoredc.com/) is providing free access to its platform to support non-profit COVID-19 research projects. As such, it is sharing eCRFs that were built according to the WHO CRF standard. Researchers can start their study or registry in less than an hour, ensuring they capture high-quality data to help drive the global research effort. At the time of publishing, 300 COVID studies haven been created, more than 200 are live and over 5 million data points have been captured on COVID-19 related projects. Further, Castor is supporting the largest global randomised "mega-trial" the find treatments for COVID-19.

The urgency of data sharing

In the post-mortem analysis of the Ebola outbreak response, failure to share relevant data in a timely way has been identified as one of the major hindrances to mounting an effective response. Although the outbreak was eventually contained, lack of data sharing and communication breakdowns delayed acknowledgement of the outbreak's severity and a coordinated response.

Ultimately, the deficiencies of data-sharing mechanisms during the Ebola outbreak became a catalyst for change. In September 2015, WHO held a consultation called "Developing Global Norms for Sharing Data and Results during Public Health Emergencies", where international stakeholders clearly stated that there must be timely and transparent pre-publication sharing of data and results during public health emergencies.[5] After considering the perspectives shared at this meeting, WHO released no-non-1sense recommendations for data sharing during any public health emergency. In fact, they urged "a paradigm shift in the approach to information sharing in emergencies, from one limited by embargoes set for publication timelines, to open sharing using modern fit-for-purpose pre-publication platforms". Of course, such a massive shift requires buy-in from researchers, journals, funders, and others.

WHO went on to prescribe open data sharing as the default response to a public health emergency, declaring that sharing results should be standard practice during a public health emergency. Their recommendations included strong encouragement of sharing epidemiological and population-based data. They warned of the great risks associated in withholding data and results arising from analyses of that data. As we have seen with COVID-19, the risks associated with epidemics are often not shouldered by a single community or nation but rather by the whole planet. The price of data hoarding is simply too high to be allowed any longer.

In line with its own recommendations, WHO immediately began working with its own networks of researchers and other experts to coordinate global work on surveillance, epidemiology, modelling, diagnostics, clinical care, and treatment of COVID-19. It also launched a Global 2019-nCoV Clinical Data Platform.[6] This allows member states to contribute anonymised clinical data, widening the breadth and depth of data collected. Data sharing also went well beyond WHO's own platform. For example, the release of full viral genome sequences through open databases resulted in the development of rapid and reliable diagnostic tests within weeks.[7]

Another issue that was brought forward during the meeting was the unacceptability of non-disclosure of clinical trial data related to research and development in the context of public health emergencies. During a health crisis, decision-makers rely on information disseminated through peer-reviewed journals and accompanying online data sets. Outside of public health emergencies, 12 months is generally considered an acceptable time frame from study completion to public disclosure. However, in an emergency context, WHO recommended that this time frame should be greatly shortened.

Medical journals have responded by taking bold steps to make information available right away. Some of these steps were first tested out during the Zika virus epidemic, when relevant manuscripts were posted online in open collections within 24 hours of submission while undergoing peer review.[8] This trend has continued during the COVID-19 outbreak. In their position statement regarding sharing data during a public health emergency, *The New England Journal of Medicine* states: "Funder signatories will require researchers undertaking work relevant to public health emergencies to establish mechanisms to share quality-assured interim and final data as rapidly and widely as possible, including with public health and research communities and the World Health Organization."[9] Additionally, most major medical journals are providing free access to any and all articles relevant to COVID-19. These changes represent a major shift away from waiting many months, even years, before making highly relevant data accessible to all interested parties.

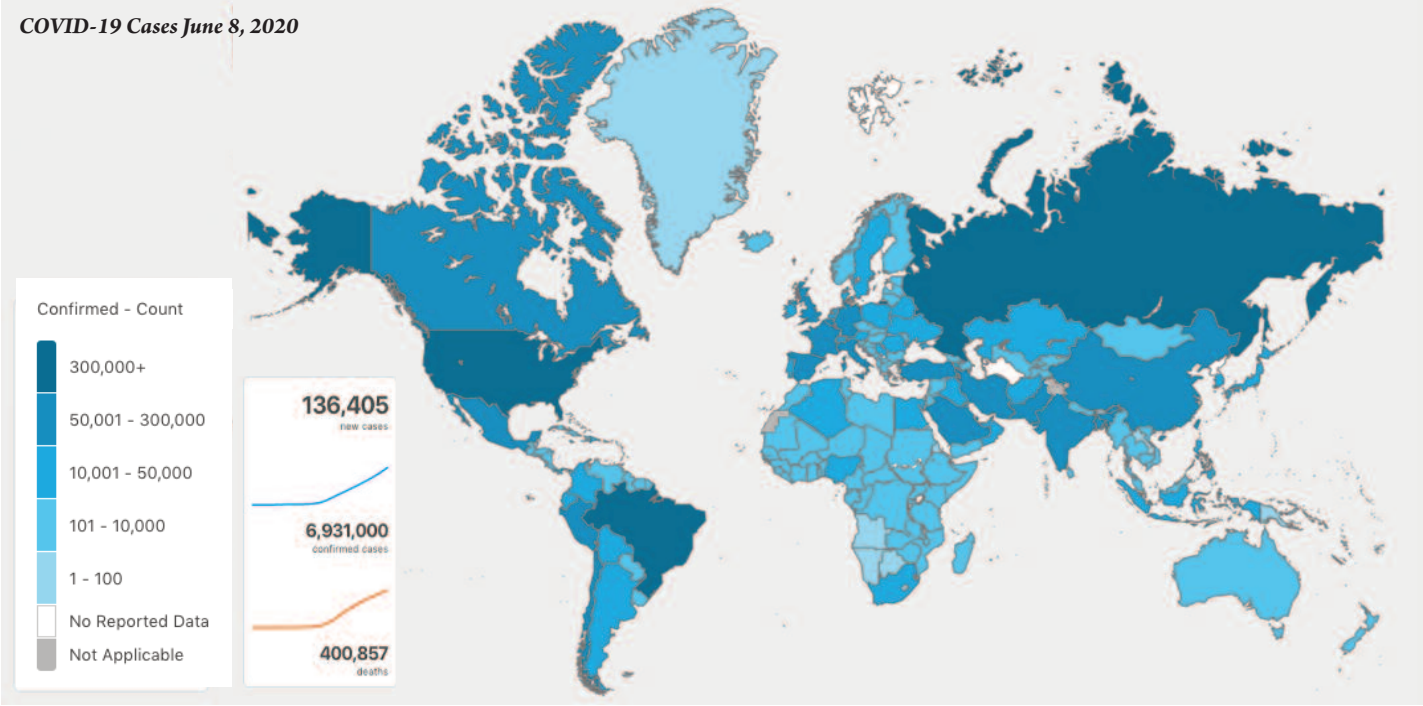
> Epidemics and pandemics spread fast. They do not wait for clinical trials or academic journals to publish results.

Conclusion

Epidemics and pandemics spread fast. They do not wait for clinical trials or academic journals to publish results. In order to contain the current crisis, our scientific communities must leverage the power of data through standardised datasets. With this latest public health emergency, we have an opportunity to get it right. We can accelerate the discovery of cures through cooperation and collaboration. The best way to save lives is to share meaningful data in realtime.

Conflicts of interest

Derk Arts is the CEO of Castor EDC, a health tech company that produces medical data capture software.



References

1. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA. 2020;323(13):1239–42.
2. World Health Organization. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). [cited 2020 Mar 2]. Available from: https://www.who.int/news-room/detail/30–01–2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)
3. Georgetown University Medical Center. Data Sharing during the West Africa Ebola Public Health Emergency: Case Study Report. [cited 2020 Mar 8]. Available from: http://www.glopid-r.org/wp-content/ uploads/2019/07/data-sharing-during-west-africa-ebola-public-health-emergency-case-study-report-georgetown.pdf
4. World Health Organization. Novel coronavirus (COVID-19) – rapid version. [cited 2020 Mar 2]. Available from: https://www.who.int/docs/default-source/coronaviruse/who-ncov-crf.pdf?sfvrsn=84766e69_2
5. World Health Organization. Developing global norms for sharing data and results during public health emergencies [cited 2020 Mar 8]. Available from: https://www.who.int/medicines/ebola-treatment/blueprint_phe_data-share-results/en/
6. World Health Organization. Coronavirus disease (COVID-19) technical guidance: The Unity Studies: Early Investigations Protocols. [cited 2020 Mar 8]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/early-investigations
7. Sheridan C. Coronavirus and the race to distribute reliable diagnostics. Nature Biotechnol. 2020;38:382–4.
8. Dye C, Bartolomeos K, Moorthy V, Kieny MP. Data sharing in public health emergencies: a call to researchers. Bull World Health Org. 2106;94:158.
9. New England Journal of Medicine. Zika Virus. [cited 2020 Mar 8]. Available from: https://www.nejm.org/zika-virus

Author information

Derk Arts, PhD, MD, is an epidemiologist and the founder of Castor. As CEO, he's responsible for Castor's growth by staying one step ahead of the field, defining long-term strategy, and building an awesome team.

Lauren McGill is a Canadian freelance medical writer with more than 15 years in various healthcare roles. She believes that storytelling can turn data into compelling narratives that promote health literacy and provider buy-in.