Tackling the next epidemic: data technology to the rescue

Dylan George[1], Tara O'Toole[2]

"*We have the tools to bring the right interventions, to the right people, in the right places to save lives. Now, we just need the will.*"
- Susan Desmond-Hellman, Science, 19 August 2016, 353(6301): p731.

Summary: We live in an era of increasingly frequent and impactful infectious disease outbreaks. Naturally occurring outbreaks will have significant regional and international security implications for the foreseeable future, given the negative impacts (i.e., death, societal disruption, and economic costs) of such events. We also live in an information era. Integrating novel and available data technologies into public health practice will improve situational awareness, help shape outbreak interventions more precisely, facilitate faster and more efficient response activities, and save lives. To realize these efficiencies, federal, state and local public health agencies need a fundamentally more aggressive and systematic adoption, use and coordination of data technologies to provide essential information for tailoring interventions during an outbreak. Current and emerging data technologies can help tackle the next epidemic.

Epidemics threaten National Security

Infectious disease outbreaks are major national security threats because they can cause high mortality in civilian populations, economic losses, social disruption, political destabilization, and reduced military effectiveness. The frequency and impact of outbreaks are increasing[3]. Epidemics will become recurring events requiring engagement at the highest levels within the U.S. government, private sector, and among the public. Recent outbreaks have stretched the response capacity of the U.S. and diverted significant amount of resources. Outbreaks also impact the private sector. Globally, pandemics will continue to negatively impact private sector economies with a conservative estimated loss of $60 billion per year[4]. Outbreaks cause a "fog of war" that create confusion among response leaders and the general public because of a lack of vital information on how to characterize risks and respond appropriately to save lives, minimize economic impact, and mitigate social disruption. These adverse consequences of epidemics could be greatly reduced by the proper use of technologies for data collection, analysis and visualization.

> *"Better data and analyses are critical to identifying, crafting and communicating the purpose and effectiveness of non-pharmaceutical interventions."*

[1] B.Next, In-Q-Tel
[2] Senior Fellow; Executive Vice President, In-Q-Tel
[3] Jones et al. 2008. Global trends in emerging infectious diseases. *Nature* **451**, 990-993 (21 February 2008) | doi:10.1038/nature06536; Received 2 August 2007; Accepted 11 December 2007
[4] 2016. The Neglected Dimension of Global Security: A Framework to Counter Infectious Disease Crises; National Academy of Medicine Report on the Global Health Risk Framework

Epidemic management demands a flow of rapid, accurate, actionable information

Typically, outbreak response activities focus on medical countermeasures (i.e., vaccines, therapeutics, diagnostics) to treat the ill and to protect the well against infection. Unfortunately, medical countermeasures have not been, and for the near-term will not be, available soon enough to mitigate outbreaks involving emerging or re-emerging pathogens as evidenced in recent outbreaks (e.g., SARS, MERS, Ebola, Zika). Over the long term, the availability of effective medical countermeasures is essential. Given today's technologies, our collective ability globally to design, develop, manufacture and deploy medical countermeasures at scale is not sufficient for providing vaccines and drugs in a timely fashion to counter ongoing outbreaks of emerging or many re-emerging pathogens.

It may be possible to develop a vaccine that would be useful in the *next* epidemic, but, mitigating a current outbreak of a novel pathogen will depend entirely on *non-pharmaceutical interventions*. These include such measures as isolation of infectious persons; hygiene measures like hand washing or covering one's mouth when coughing; avoiding crowded spaces, travel advisories and other actions intended to break the chain of transmission. Such interventions, especially social distancing and safe burial practices, were the main means adopted by West African communities that stopped the Ebola epidemic. Indeed, reliance on non-pharmaceutical actions has been the predominant mode of epidemic management throughout history, even in modern times. If vaccines are not available at the initial stages of an outbreak, non-pharmaceutical interventions are the *only* available means of mitigating the outbreak.

Data are critical for rapid and effective responses, but these data have not been readily available nor timely. Better data and analyses are critical to identifying, crafting and communicating the purpose and effectiveness of non-pharmaceutical interventions. Ron Klain, the Ebola Coordinator for the U.S. National Security Council emphatically stated that the epidemiological data (i.e. case counts by location through time) were not as helpful in shaping the response as one might hope. Klain noted, "when I assumed my role, the data then available were not what we needed to guide decisions"[5]. The lamentable reality is that the lack of timely, reliable data left many resource allocation decisions in the realm of guesswork. Application of the right data to the response effort could have improved response effectiveness.

> *"The key to effective decision-making and the use of non-pharmaceutical interventions is obtaining the **essential information** needed to gain an accurate understanding of the situation on the ground among the affected population…"*

Essential data are needed for outbreak management

The "right data" needed for outbreak management have historically been collected by astute clinicians and by public health officials conducting "shoe leather" epidemiology (e.g. outbreak investigations, contact tracing, etc.). Such experts remain critically important in detecting outbreaks, characterizing risks, and implementing mitigation actions.

---

[5] Personal communication at a B.Next workshop on key decisions for outbreak management

Resource and logistical constraints limit the scalability of such expert effort, however. In the U.S., over 3000 state and local public health departments are responsible for outbreak response (and all other public health tasks). If something suspicious is discovered by public health officials, physicians, or via laboratory results, the findings are, in theory, reported from the locality to state officials and then to the federal Centers for Disease Control and Prevention (CDC). The CDC may then alert the World Health Organization. This fragmented, distributed and variable public health system is slow to mobilize and difficult to organize due to the number of organizations and layers involved.

The capacity to deploy trained epidemiologists during large, possibly multiple outbreaks will be operationally challenging and in some situations may outstrip the capacity of the Federal government to support state and local public health agencies. Digital technologies could serve as a force multiplier and supply some solutions to this dilemma. Even limited application of digital technologies to outbreak response could make a big difference.

The key to effective decision-making and the use of non-pharmaceutical interventions is obtaining the *essential information* needed to gain an accurate understanding of the situation on the ground among the affected population (i.e. how is the disease spread, how clinically severe is the disease, how many are ill, where are they?) and information about the nature of the pathogen causing the outbreak and its effects on hosts. Of particular importance, is information about the clinical *severity* of the infection - e.g. does infection cause a high fatality rate among all or some of those infected? – and its *transmissibility*. For example, how easily and rapidly is the infection transmitted, how might local conditions and practices increase or discourage spread of infection, and which types of interventions effectively stop transmission. Some aspects of these essential data may be available from past experience or research. But the frequency of emerging infectious disease outbreaks globally is increasing[3], and typically we know little about the pathogens responsible or the nature of the illness they cause (e.g., 2015-2016 Zika virus epidemic).

> *"We need a fundamental reconsideration of how to use combinations of data technologies for effective response management."*

Acquiring the disparate data needed to provide this information is a huge task with significant time pressures. Technologies that help collect, clean, analyze, visualize and communicate the meaning of these data must become a priority of epidemic management if we are to become faster and more effective at responding to outbreaks.

An information supply chain is needed for epidemic management

Data technologies have revolutionized modern life. Data powers the collaborative economy and informs countless daily activities through the use of mobile devices, crowd-sourcing efforts, geo-location capabilities, and even inexpensive genomic sequencing. We are awash in data that are being collected, shared and analyzed at speeds and scales which were unimaginable even a decade ago. In contrast, data technologies have ***not*** revolutionized healthcare and outbreak response capabilities. Technological advances have been marginally and sporadically supported, uncoordinated in developing capabilities, and haphazardly applied to outbreak management. For example, pen and paper remain the main means of record keeping in many clinical settings in low-resource settings. Given how transformative the use of data has been in so many areas, it is striking to realize that public health practice lacks reliable systems to collect, curate, analyze and communicate data especially during public health emergencies. Also, data technologies are expected to be transformational in healthcare in the coming years. We need to apply digital technologies to outbreak management if we are to become more effective at outbreak response.

Incorporating advances in data technologies into public health practice could help leaders make informed decisions about epidemic management and significantly improve outbreak response. The fundamental workflow from data collection to decision and communication can be considered the "information supply chain" for outbreak detection and response (Figure 1). This basic workflow concept applies to many data-dependent activities. While simplistic, it is a useful metaphor for guiding applications of appropriate technologies at each step in the chain. Systematically adopting existing and emerging technological advances to each segment of the supply chain could eventually lead to the information flow across the entire supply chain becoming more reliable, efficient and timely. A less ad hoc, more coherent and reliable information supply chain would in turn improve overall epidemic preparedness and response.

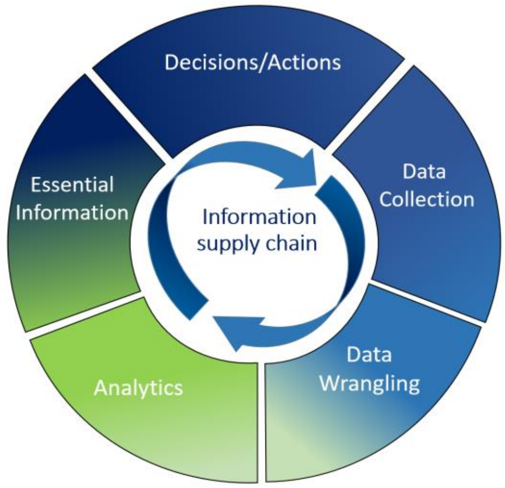

Figure 1: An information supply chain for outbreak management

Establishing an information supply chain for epidemic response – examples of enabling technologies

To be successful the information supply chain will need to be comprised of a range of data, curating technologies, and analytics that are rooted in providing the essential information for decision-making. Many challenges with data collection, curating, analytics and communication could be successfully and elegantly addressed by current and emerging technological solutions. The authors are not advocating adoption of a particular technology. Rather, we offer the following examples of specific data technologies to illustrate that helpful information technologies are available which could greatly improve the information supply chain for outbreak management. Below are examples of technologies that if adopted would facilitate response efforts.

Novel data or means of collection: New data sources relevant to epidemic management are proliferating[6]. Social media, call data records, patterns of mobile phone use, geo-location information, satellite imagery and genomics are transforming the notion of public health data, and could greatly aid epidemic response.

[6] Hay SI, George DB, Moyes CL, Brownstein JS (2013) Big Data Opportunities for Global Infectious Disease Surveillance. PLoS Med 10(4): e1001413. doi:10.1371/journal.pmed.1001413

Crowdsourcing methods have been used to collect and validate information on outbreaks. ProMED[7] (Program for Monitoring Emerging Diseases) which is hosted by the International Society for Infectious Diseases offers an example of the benefits of digital technologies. ProMED was begun by volunteer doctors as a non-profit organization that employs veterinarians and clinicians around the world who "monitor" and then post reports of infectious disease cases and outbreak information. Anyone can report instances of unusual or suspicious disease. ProMed was the first to alert the world about the SARS epidemic and has significantly contributed to reducing the time to detect an outbreak. ProMED joined forces with HealthMap[8], based out of John Brownstein's computational epidemiology lab at Boston Children's Hospital, to visualize these data and add scraped news stories from the internet as early indicators. These sources of data have been used broadly among those monitoring outbreaks.

Of particular interest is the use of *genomic sequences for outbreak management*. For example, sequencing the genomes of the pathogens responsible for disease outbreaks has become essential to outbreak investigations over the last decade. However, until recently, analyses using genomic sequences have been largely retrospective. Today, because of ongoing improvements in the speed and accuracy of sequencing technologies and parallel decreases in the cost to do so, genomic sequencing of new and emerging pathogens is beginning to become routine in near real-time. As the cost of sequencing genomes continues to drop and the portability of field sequencing capabilities increases, genomic sequencing will allow faster, more precise characterization of outbreak risks. The data generated from near real-time sequencing can be used for phylogenetic analyses (i.e., making family trees of how the pathogens relate to one another). Such information can be used to determine how a pathogen is spreading, inform contact tracing, locate hot spots of transmission, and to detect adaptive evolution of the pathogen that might require changes in public health actions.

During the 2013-15 Ebola outbreak[9], and the ongoing Zika crisis, academics used field-portable sequencing devices to improve scientific understanding. But these efforts were *ad hoc* and uncoordinated, and the data generated did not flow smoothly to the greater scientific and public health community or to key decision-makers. Additionally, the software for ingesting, cleaning, analyzing and visualizing sequence data were developed largely through the generous efforts of academic researchers[10]. Developing an "enterprise software" that allows larger numbers of users to both enter and analyze relevant outbreak genomic data without bespoke academic support and involvement could significantly benefit outbreak management.

The need for more robust bioinformatics software is especially acute because cheap, field-ready, sequencing technologies will significantly increase the quantity and availability of near-real time sequencing data during an outbreak. Once sequencing becomes ubiquitous, there will be a pressing need to ingest and make sense of the huge information flows from genomic sequence data.

Data cleansing: Extracting, transforming, and loading data, or rather data curation and cleansing, into usable formats is a common problem across many fields, and is no exception during an

---

[7] http://www.promedmail.org

[8] http://www.healthmap.org/

[9] Quick et al. 2016. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530, 228–232 (11 February 2016) doi:10.1038/nature16996

[10] Gire et al. Science. 2014 Sep 12;345(6202):1369-72. doi: 10.1126/science.1259657. Epub 2014 Aug 28. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak.

outbreak. Arguably, data cleansing is one of the least fun aspects of many data science fields. During the Ebola outbreak in West Africa, a significant portion of the local data collection was accomplished with pencil on paper. Data on a single patient was often collected by multiple sources and in several different locations (e.g. treatment units, diagnostic labs, contact tracing teams, etc.). The same person's name was often spelled several ways. Before these data could be analyzed, a large team was needed to manually collate and cleanse the data in order to disambiguate records.

During the Ebola response, digitizing the data into machine readable formats was done manually and delayed analysis of key epidemiological information. Recently, advances in a range of technologies such as machine learning techniques and crowdsourcing (i.e., Amazon Mechanical Turk) have been applied to the problem of digitizing non-machine readable text with considerable success. Applications of such technologies have the potential to greatly speed the collection and wrangling of hand-written data, and if they had been used during the Ebola response, they might have allowed for faster data curation and cleansing. Additionally, a number of the individuals that were manually collating and cleansing the data could have been released to instead apply their epidemiology and analytical expertise elsewhere.

Analytics: The amount, availability and variety of digital data is increasing exponentially. These caches of big data have given rise to opportunities to use these data for detecting outbreaks, characterizing their risks, projecting their impact and guiding response activities. Analytics have become a common feature in many parts of modern living. For example, predictive analytics, forecasting, machine learning, recommendation engines, geospatial analytics, and even phylogenetics are becoming common place. Many of these analytical approaches have not been applied consistently, effectively, robustly or commonly in public health or to outbreak management. More frequent use of new analytic techniques could improve outbreak management.

Advances in infectious disease forecasting of ongoing outbreaks could significantly improve decision-making. For example, during the Ebola outbreak, the U.S. government developed forecasts to anticipate challenges associated with clinical vaccine trials. The results informed adjustments to trial designs that more accurately considered changes in incidence levels as the outbreak was waning.

Other analytical products have used social media data and predictive analytic techniques to anticipate when and where cold and flu symptoms would be most prevalent so that pharmacies could pre-emptively move cold and flu medications to those locations. In the near future, predictive analytics will be more broadly applied to facilitate operationally relevant decision-making during unique and seasonal outbreaks (e.g., influenza, dengue).

Another area of promising analytics is geospatial analyses. This type of analysis is enabled by advances in mobile technologies that geocode activities relevant to outbreaks. These analytics have been used effectively to provide assessments of population movements[11], locations of high risk groups [12], locations of clinics, geographies with high potential for disease spillover from animal populations, and general high risk geographies[13,14]. For example, during

---

[11] Deville et al . 2014. Dynamic population mapping using mobile phone data.PNAS.www.pnas.org/cgi/doi/10.1073/pnas.1408439111

[12] Bharti et al 2015. International Health. Int. Health (2015) 7 (2): 90-98.doi: 10.1093/inthealth/ihv003. Remotely measuring populations during a crisis by overlaying two data sources

[13] Hay SI, George DB, Moyes CL, Brownstein JS (2013) Big Data Opportunities for Global Infectious Disease Surveillance. PLoS Med 10(4): e1001413. doi:10.1371/journal.pmed.1001413

[14] Pigott et al. 2016. Updates to the zoonotic niche map of Ebola virus disease in Africa. eLife 2016;5:e16412. DOI:

the Zika outbreak, areas of high risk for transmission of the virus were quickly defined in the United States and internationally[15]. These analyses were translated into travel advisories and guidance for protecting oneself against infection[16]. Using a range of existing and emerging technologies to improve our understanding of where and when outbreaks occur has been exceptionally useful and will continue to be helpful in future infectious disease outbreaks.

Visualization: Providing decision makers with more accurate, compelling and timely visualizations of outbreak data could significantly improve epidemic management. Currently, public health analysts are limited to inefficient applications to quickly visualize and collaboratively analyze data. Visualizations can be exceptionally challenging and time consuming to create, often requiring more time than the actual analyses. Technologies that can shorten this time will be valuable in outbreak management, and web-based visualization platforms that enable collaborative data science workflows appear promising.

Conclusion

Outbreaks continue to threaten national security. When it comes to outbreak responses time matters because it means lives and livelihoods. Improving response times for activities that have proven to be effective (i.e, non-pharamceutical interventions) need to be prioritized. The full potential of surveillance and advanced analytics for improving outbreak management has not yet been realized and, unfortunately, is not yet adequate to the task. We need a fundamental reconsideration of how to use combinations of data technologies for effective response management. Accomplishing this reconsideration and implementing it effectively will allow for faster, better, stronger responses. Past outbreaks have threatened national security, but they do not need to be as significant a threat in the future. Current and emerging data technologies can help tackle the next epidemic.

Acknowledgements: We thank Lisa Porter, Stephanie Rogers, Kevin O'Connell, Tiara Jones, and Joe Buccina for insightful discussions and helpful suggestions on early drafts.

http://dx.doi.org/10.7554/eLife.16412

[15] *Bogoch* et al. 2016. Anticipating the international spread of Zika virus from Brazil. Lancet. January 14, 2016 http://dx.doi.org/10.1016/ S0140-6736(16)00080-5; and, Messina et al 2016. Mapping global environmental suitability for Zika virus. DOI: http://dx.doi.org/10.7554/eLife.15272. April 19, 2016. eLife 2016;5:e15272

[16] https://www.nytimes.com/2016/01/14/health/zika-virus-cdc-travel-warning-brazil-caribbean.html

