

Tests, Surgical Masks, Hospital Beds, and Ventilators: Add Big Data to the List of Tools to Fight the Coronavirus That Are in Short Supply

Dennis P. Scanlon, PhD, MA; and Mark B. Stephens, MD, MS

A Tool for the Times

Let's be very clear. Health care professionals around the world, such as doctors, nurses, respiratory therapists, pharmacists, aides, and public and private health administrators, are heroically and intentionally engaged in warfare against a lethal virus: severe acute respiratory syndrome coronavirus 2, which causes coronavirus disease 2019 (COVID-19). Various tools to fight the opponent—COVID-19 testing and surgical masks, as well as more structural resources such as intensive care unit beds and ventilators—are well documented to be in short supply for the daunting challenge. Some have suggested that what we are requiring health care workers to do is analogous to sending soldiers into combat without basic tools of the trade, such as ammunition and body armor. Add 1 more thing to the list of resources that are unavailable or not being effectively mobilized in the current battle: big data.¹

This commentary discusses what “big data” means and why it matters in this particular pandemic, but the bottom line is that big data is used for critical decision making in the military and just about any other coordinated effort in our society, be it economic strategy or the operation of humanitarian efforts. The use of big data has been woefully absent in our nation's response to the COVID-19 pandemic, and in public health and health policy planning more generally, despite billions spent by the US government since the passage of the Health Information Technology for Economic and Clinical Health (HITECH) Act in 2009. Ideally, soldiers go into combat informed about the opponent's location, tactics, strengths, and weaknesses to the maximum degree possible from existing operational intelligence. Our doctors, health care workers, public health specialists, and state and local officials are fighting this battle not only without equipment, but sadly also without the kind of central intelligence that could be gleaned from various types of big data analyses, including knowing for sure where the opponent is and how it is moving. Intelligence of this sort cannot, alone, solve the problem or replace the Herculean efforts of health professionals on the front lines, but it certainly could help.

What Is Big Data? What Can It Do?

What is big data and why might it make a difference in the current pandemic? Big data means 2 things: real-time availability of current and historical data from multiple sources and the ability to analyze and process those data in real time. Right now, big data could help identify potential clues about the immediate (and future) impact of COVID-19—for example, by providing a clearer understanding about those who may be most at risk for the worst outcomes or helping to describe how providers have been treating patients with COVID-19 symptoms, even if tests confirming positive cases have been absent. For instance, data could potentially be used to identify individuals who were infected and then recovered uneventfully; this would help inform development, testing, and administration of new antibody-directed therapies,² as well as potentially segmented social distancing policies.

Big data is not a panacea. Its potential lies in using information about thousands of cases, from patients within local health systems to the state, federal, and international levels. The goal is to look for patterns—for example, what are the characteristics of those who seem to tolerate the virus, exhibit minimal symptoms, and recover without resultant health conditions? Are there patterns in the case histories, well before infection, of affected patients who progress to extreme illness and require critical care? Could these patterns be used to better allocate and move scarce health care resources to the areas next predicted to have the most need?

Aid to the Front Lines

Doctors and medical staff at the patient's bedside continually use data to inform their decisions. Their focus, however, is primarily on the individual patient in front of them at any moment in time: an “N of 1.” Although individual clinicians may discern local patterns and discuss common characteristics with colleagues, providers are not trained to gather or sort through hundreds or thousands of cases to look for patterns and clues. Nor should providers be expected to do this, because their primary role is to make critical treatment decisions at the bedside, affording each patient the respect of being an “N of 1.” Others who are highly trained in this type of analysis

can support the fight by delivering critical information for clinical and scientific consideration to the front lines. Although not a cure, these data can inform prevention and treatment strategies, patient risk segmentation, and approaches to concepts such as social distancing to better account for portions of the population who are at particularly high risk.

To help illustrate the point, in the pandemic thus far, experts have repeatedly referred to data suggesting that those most at risk are elderly patients and/or those with underlying health conditions. Stories continue to emerge, however, about younger individuals and those without any previously documented health conditions who have serious disease. Although older age, male gender, and underlying conditions (eg, heart disease, hypertension, diabetes, history of smoking, pneumonia, asthma) are lumped into the broad category of the population most at risk, it is apparent that individuals with these characteristics do not represent a homogenous group. Instead, individuals can be further sorted into subgroups to capture specific types and numbers of conditions, length of time since diagnosis of condition, current medications, and tobacco or vaping use, to name a few.

Identifying Patterns for Clues

If thousands of cases are examined, more patterns may emerge and yield data ripe for analysis. Examples could include blood type, genotype, recent presentation and treatment of symptoms related to the flu or other respiratory illnesses, and number of office visits and hospitalizations. As an example, a highly reported media account of community spread has been the tragic circumstance of 4 individuals from the same family dying of COVID-19, possibly passing the infection at a family dinner gathering.³ In another widely reported account, a *New York Times* article⁴ described early community spread in Seattle that originated within a group of friends in their 40s who were also at a dinner party. In that case, many who appeared asymptomatic at the party fell sick several days later and only became aware they had contracted COVID-19 when 1 partygoer was tested⁵ by the University of Washington's laboratory conducting the Seattle Flu Study. Early on, COVID-19 testing was controversial for those who did not present as high-risk (ie, an otherwise healthy 40-something), but that move informed the patient about her infection and prompted the subsequently symptomatic party attendees to seek testing; the COVID-19–positive friends spread the virus to people who were not at the dinner party.

A logical question is what made the spread in the family group deadly compared with the spread in the group of friends? Could it be the presence of specific health conditions, genetic susceptibility, interactions with other concurrent symptoms or medications, age, or some combination of these? Big data could potentially help to tease out what makes the virus so aggressive or even lethal in some circumstances but milder in others, offering clues that could allow doctors and policy makers to tailor treatment approaches and health policy interventions more effectively.

Challenges to Data Collection and Analysis

Why not “just do it,” then, one might ask? The short answer is that real-time data and/or readily available analytic expertise with the infrastructure and permissions to use those data are lacking. Current data are at the aggregate level and very general—patient age, presence of underlying condition, and death. Reports of patient recovery seem to be taken as a singularly uniform experience across all infected patients. At the treatment level, more details, including specific information regarding blood oxygen levels, respiratory function, medications, and doses administered for hospitalized patients, would be extremely helpful. Patients' historical medical records can also provide helpful data (eg, influenza vaccination status, presence of fever or upper respiratory symptoms in the weeks or months prior to infection). In the United States, where the absence of early and ongoing testing has been a public health failure of the highest magnitude, clues from ambulatory visits of symptomatic patients in the first 8 weeks could give a better sense of the true denominator of those infected.

On March 12, 2020,⁶ Ohio Governor Mike DeWine publicly announced that the Ohio Department of Health estimated that more than 100,000 individuals, or 1% of the state's population,⁷ were likely infected with COVID-19. If accurate, such estimates could be viewed as encouraging, as the reported number of COVID-19 deaths in Ohio at the time was very low, which might indicate that many in the population had tolerated the infection well and potentially had even recovered and built up immunity. Unfortunately, in the absence of widespread testing, such estimations are merely best guesses. To the extent that some patients included in available estimates experienced symptoms that resulted in a health care visit (ie, for symptoms such as fever, bronchitis, pneumonia, influenza, or respiratory distress), an analysis of recent historical data might prove helpful in narrowing prevalence estimates.

HITECH Act: Real-time Data Collection and Analysis

Currently, no mechanism exists for central collection of such data. Instead, data collection primarily occurs within individual patient records kept by health care providers or, increasingly, within large health care systems. Sadly, relatively few of these providers or systems (outside of the largest and most advanced systems) have the capacity to routinely and systematically query, analyze, and learn from potential big data patterns in real time.

The HITECH Act was passed as part of the American Recovery and Reinvestment Act of 2009. This federal stimulus package focused on boosting the US economy after the 2008 recession. To date, HITECH has spent \$36 billion in federal funds for electronic health record (EHR) implementation and associated elements such as interoperable information sharing across data platforms, clinical decision support, and meaningful use of data to improve patient care. The initial vision for meaningful use was to give providers access to real-time data to improve clinical decision making at the point of care. After more than a decade, many believe that HITECH has not augmented⁸ the ability of health care providers and health

care systems to query and use data in real time to inform patient care.⁹ It is a tragic irony that the potentially valuable use of data, an aim born of an economic stimulus with our last recession, is not readily available in the largest fight the US health system has ever seen, as the economy slides into the next recession.

“ IT IS A TRAGIC IRONY THAT THE POTENTIALLY VALUABLE USE OF DATA IS NOT READILY AVAILABLE IN THE LARGEST FIGHT THE US HEALTH SYSTEM HAS EVER SEEN. ”

Clinical Experimentation in Complex Situations With High Uncertainty

To be certain, some of the toughest cases to treat in all of medicine are bacterial or viral infections for which known approaches or therapies either do not work or are not available. COVID-19 brings the conversation about real-time use of big data to inform clinical care into sharp focus. Clinicians and health care workers on the front lines need all the help they can get. In addition, although COVID-19 is clearly a significant and complicated virus requiring all hands on deck, clinicians routinely encounter other pathogens during the flu season when big data could help inform decisions ranging from resource utilization to anticipated timing of disease outbreaks. Big data could inform decisions about specific treatments for patients who are nonresponsive to routine therapies or patients with atypical responses to common viral, fungal, or parasitic pathogens. As information comes out about the potential roles of hydroxychloroquine and azithromycin in the treatment of COVID-19,¹⁰ for example, the systematic application of big data could sharpen the ability to provide more evidence to critical bedside decisions. Off-label use and heterogeneity in treatment approaches inevitably exist in complex situations and in situations with high degrees of uncertainty in which treatment options seem limited. That is the point at which, by necessity, therapeutic experimentation starts to occur. In real-world clinical practice, when experimentation becomes necessary, there is not random assignment. But big data analysis can still capture and analyze the impact of this treatment variation, potentially identifying important patterns and even using advanced statistical tools in an attempt to create pseudo comparison groups to better discern what may or may not help or who might be at greater or lesser risk for the most significant health consequences of the virus. Big data could also guide screening and treatment decisions for an illness that was previously circulating and relatively undetected in the community. Put simply, big data can help harvest valuable lessons from real-time clinical practice to assist those on the front lines.

Data of a Different Type to Supplement Existing Efforts

Readers should not assume that in the current fight, no data are being collected or no valuable analysis is occurring. Instead, public entities like the CDC in the United States and similar entities across the world are collecting information on infected and treated cases, as well as cases requiring significant health system and medical support. The CDC has long had a formal mechanism for tracking known diseases such as influenza.¹¹ Epidemiologists, trained in the identification, spread, and tracking of disease, use available data and develop mechanisms for the collection of new data; they are critically important in the current pandemic and regularly provide some of the most useful information to inform clinicians and public policy. Medical specialty networks, such as national and international professionals in fields such as infectious disease, have published studies about treated cases and treatment approaches.¹² What many of these approaches lack, however, is historical information on patient populations (ie, those infected and treated for significant symptoms, those potentially infected but not requiring treatment, or those who were infected and treated but were never documented because symptoms were minor or coded as presenting with other conditions). In the United States, these types of data primarily live within patient EHRs maintained largely by individual health care practices and health systems. In theory, the use of EHRs should allow these data to be utilized, but in practice, this does not consistently happen and is a missed opportunity.

Using big data does come with privacy concerns, and the US Health Insurance Portability and Accountability Act specifies rules regarding use, including the delineation between “research” versus active population health management and clinical quality improvement.¹³ Following privacy laws is essential, and there are ways to do so while allowing clinicians to learn at the same time. Unfortunately, despite HITECH and other investment in EHRs, realizing the value in big data is for the most part missing in the current COVID-19 crisis, a lesson that must be carefully considered once the acute phase of the pandemic slows.

Clinical and Policy Implications

Why does all this matter? For 2 reasons. First, the ability to access and use patient-level data from clinical populations could potentially help with the ongoing operational challenges of the immediate clinical response. Second, this type of data could help inform decisions regarding not only health but also policies that affect the economy, education, and other aspects of societal welfare. A growing field in the last decades has recognized the critical importance of social determinants of health¹⁴—nontraditional health and medical care services and resources, such as income, housing, food security, transportation, and social isolation, and their enormous impact on the overall health of individuals and populations. All policy decisions currently being made have short- and long-term impacts on the health of individuals, communities, states, and the nation. Big data could inform which policies are implemented, while concurrently informing the health impacts

on the back end. Big data involve not only individual-level health data in real time, but also the ability to link information about social programs, such as Medicaid, housing, criminal justice, child and family welfare, and means-tested subsidy programs. Appropriate linkages of these types of data, attending to relevant privacy concerns, could occur at the state level as integrated data systems¹⁵ to offer the ability to estimate the impact of various federal and state efforts to provide services and aid. Although our work in the Commonwealth of Pennsylvania has yielded more awareness of the potential value of such integrated data, it is unfortunate that few states currently have fully integrated data systems, which will challenge the tracking and assessment of the impact of the policy response to COVID-19.

Current Demands and Long-term Thinking

In the fight against the COVID-19 virus, many critical shortages have emerged. Clearly, providing for the frontline needs of health care workers and patients should be the nation's primary focus. Again, big data can help deliver intelligence to those front lines so the providers can be strategic at each bedside. Once the immediate combat wanes, big data will be necessary to evaluate the success of various policy and spending decisions made during the crisis, as well as to help estimate the broader impact that the economic consequences of COVID-19 have had on health and its social determinants. Finally, an important future discussion to have is why, after more than a decade and tens of billions of dollars of investment, many of the nation's providers and health systems are not better equipped to use valuable patient- and population-level data in real time to inform important treatment and management decisions. ■

Author Affiliations: Department of Health Policy & Administration and Center for Health Care and Policy Research (DPS) and College of Medicine (MBS), The Pennsylvania State University, University Park, PA.

Source of Funding: None.

Author Disclosures: The authors report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

Authorship Information: Concept and design (DPS, MBS); analysis and interpretation of data (MBS); drafting of the manuscript (DPS, MBS); and critical revision of the manuscript for important intellectual content (DPS, MBS).

Address Correspondence to: Dennis P. Scanlon, PhD, Department of Health Policy & Administration and Center for Health Care Policy Research, The Pennsylvania State University, 504 Ford Bldg, University Park, PA 16802-6500. Email: dpscanlon@psu.edu.

REFERENCES

1. Lohr S. The age of big data. *New York Times*. February 11, 2012. Accessed April 3, 2020. <https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html>
2. Palca J. FDA expedites treatment of seriously ill COVID-19 patients with experimental plasma. NPR. March 24, 2020. Accessed April 3, 2020. <https://www.npr.org/sections/coronavirus-live-updates/2020/03/24/820939536/fda-expedites-treatment-of-seriously-ill-covid-19-patients-with-experimental-pla>
3. Tully T. Coronavirus ravages 7 members of a single family, killing 4. *New York Times*. March 18, 2020. Accessed April 3, 2020. <https://www.nytimes.com/2020/03/18/nyregion/new-jersey-family-coronavirus.html>
4. Sanders L. Partygoers in Seattle, a suburban D.C. mom: coronavirus in the community. *New York Times*. March 18, 2020. Accessed April 3, 2020. <https://www.nytimes.com/2020/03/18/well/live/coronavirus-symptoms-diagnosis-covid-19-community.html>
5. Phan S. Seattle woman shares her coronavirus experience. KOMO News. March 11, 2020. Accessed April 3, 2020. <https://komonews.com/news/coronavirus/seattle-woman-shares-her-coronavirus-experience>
6. @GovMikeDeWine. @DrAmyActon: I know it is hard to understand #COVID19 since we can't see it, but we know that 1% of our population is carrying this virus today – that's over 100,000 people. March 12, 2020. Accessed April 3, 2020. <https://twitter.com/GovMikeDeWine/status/1238177953126604801>
7. Sullivan P. Ohio health official estimates 100,000 people in state have coronavirus. *The Hill*. March 12, 2020. Accessed April 3, 2020. <https://thehill.com/policy/healthcare/487329-ohio-health-official-estimates-100000-people-in-state-have-coronavirus>
8. Thune J, Alexander L, Roberts P, Burr R, Enzi M. Where is HITECH's \$35 billion dollar investment going? *Health Affairs*. March 4, 2015. Accessed April 3, 2020. <https://www.healthaffairs.org/doi/10.1377/hlthlog.20150304.045199/full/>
9. Reisman M. EHRs: the challenge of making electronic data usable and interoperable. *P T*. 2017;42(9):572-575.
10. Information for clinicians on therapeutic options for patients with COVID-19. CDC. March 21, 2020. Updated April 7, 2020. Accessed April 8, 2020. <https://www.cdc.gov/coronavirus/2019-ncov/hcp/therapeutic-options.html>
11. Flu activity & surveillance. CDC. October 11, 2019. Accessed April 3, 2020. <https://www.cdc.gov/flu/weekly/fluactivitysurv.htm>
12. Holshue ML, DeBolt C, Lindquist S, et al. First case of 2019 novel coronavirus in the United States. *N Engl J Med*. 2020;382(10):929-936. doi: 10.1056/NEJMoa2001191
13. Gregory KE. Differentiating between research and quality improvement. *J Perinat Neonatal Nurs*. 2015;29(2):100-102. doi: 10.1097/JPN.0000000000000107
14. Social determinants of health. Healthy People. Accessed April 3, 2020. <https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health>
15. Integrated Data Systems (IDS). Actionable Intelligence for Social Policy. Accessed April 3, 2020. <https://www.aisp.upenn.edu/integrated-data-systems/>

Visit ajmc.com/link/4587 to download PDF