

The ubiquity of data & communication: A double-edged sword for disparities

Robert M. Califf

abstract

The fourth industrial revolution, in which most information is stored in digital form, is characterized by connectivity and communication among people and groups via, for instance, cell phones and smart watches. The amount of information now generated about people's health-related activities is multiple log orders more voluminous and complex than the data currently captured in the electronic health record from patient interactions with clinicians. Despite the data's complexity, it is now possible for health care administrators, policymakers, and clinical researchers to develop—and then test—data-informed interventions that could reduce health disparities. For example, programs initiated by a county government and a major medical system have, respectively, improved asthma management and reduced lead exposure in their localities. Use of big data can be a double-edged sword, however. The technology that allows for high-end use of data also opens the way to increasing disparities, as could happen, for instance, if geospatial information were used to locate clinics in places that optimize profit rather than meet health needs. Efforts are underway to limit this risk.

Core Findings

What is the issue?

The advent of the fourth industrial revolution has brought a wealth of new technologies and an exponential increase in information. This presents an unprecedented opportunity to address disparities in health care. New methods of data curation and analysis can lead to more effective interventions. But without societal involvement and participation in guiding new technology, health disparities are at risk of becoming more acute.

How can you act?

Selected recommendations include:

- 1) Developing a single standard for health care data pertinent to health outcomes
- 2) Using new technologies to provide useful information in a way that is tailored to the specific health needs of an individual, family, or population group

Who should take the lead?

Researchers, policymakers, and stakeholders in health care

For decades, it has been common knowledge that vast disparities in health outcomes and access to health care occur both in the United States and across the globe. Until recently, clinicians, health care administrators, policymakers, and clinical researchers have lacked timely access to the type and quantity of information that would enable interventions that might ameliorate disparities between individuals and populations. With the advent of the fourth industrial revolution,¹ that situation is quickly changing.

Whereas the third industrial revolution was marked by the introduction of digital technologies such as the Internet and personal computers, the fourth industrial revolution is characterized by the merging of biological, physical, and information sciences.¹ This fourth revolution has enormously expanded connectivity and communication among individuals and groups. For instance, today's cell phones reach almost everyone on the planet, regardless of income, education, and physical location, and they provide instant access to vast digital communication and information systems. Given the almost universal ability to connect to distributed cloud computing, the computing power accessible through a typical cell phone today exceeds the computational resources of entire universities or medical centers from just a few years ago.

New technologies for mapping and wayfinding illustrate the potential of the connectivity and communication that are hallmarks of the fourth revolution. Many drivers already rely on global positioning systems to help them navigate unfamiliar streets and highways, using information that is accessed, contextualized, and integrated in real time. Smart wayfinding apps provide a driver with current data about the driver's location, traffic and road conditions, upcoming businesses and landmarks, and reports from other drivers. This information helps drivers avoid accidents, potholes, and traffic jams. Drivers also receive predictions about the impact of interventions (such as the effect on driving time that taking a different route might have) to support their decisionmaking. Actual results of driving experiences are then fed back

into the algorithms to improve them. Some key innovations relate to data fluidity—the capacity for data to flow easily and without undue friction among all the users who need it—and data latency—the speed at which data, once gathered, is available for analysis.

What if a similar approach were taken to human health care, including treatment and prevention? I argue that the ubiquity of data and its rapid communication provide previously unimaginable resources for understanding and addressing health disparities. One simple possibility is that home services could be better coordinated by systematically applying optimized transportation routing and scheduling of the kind now available with smartphone apps. Thus, a patient with uncontrolled diabetes living in an underresourced neighborhood would receive more frequent home visits from a nurse than a less ill person would.

Already, in Durham County, North Carolina, a data system connected to the Duke University Health System has contributed to a substantial improvement in locating children with elevated blood lead levels.² The Duke system created a map that estimates household lead exposure risk based on county tax assessor data, blood lead screening results from clinic visits, and census data. Stakeholders, including the Durham County Health Department and several community advocacy groups, have used this map to reach at-risk families.

Similarly, in Louisville, Kentucky, the public-private AIR Louisville consortium is helping local residents manage their asthma, a disorder that disproportionately affects Black children and people living below the poverty line.³ Relying on electronic sensors in inhalers, the program provides feedback about triggers, adherence to treatment, and level of control to asthma patients, which has resulted in a 78% reduction in rescue inhaler use and a 48% improvement in the number of symptom-free days. After combining crowd-sourced data on inhaler use with environmental information, a government and community activist team crafted policy recommendations to lower the incidence of asthma attacks citywide, such as increasing the

“Researchers and clinicians are now able to amass novel kinds of biological data, such as an individual’s genetic code”

tree canopy (to reduce air pollution and urban heat), requiring that facilities with vulnerable populations (such as children and the elderly) be located at least 500 feet from roadways with high traffic and high emissions, and developing a community notification system that alerts asthma sufferers when high-risk conditions are about to occur.

Although this powerful technology can be harnessed to reduce health disparities, it may also exacerbate them.⁴ For instance, data analysis has revealed that the Russian government is using bots to spread skepticism about the safety of vaccines on Twitter in an apparent attempt to create discord in the United States.⁵ But many times, negative consequences may well happen inadvertently. *Machine learning* is a type of artificial intelligence that uses computer algorithms to predict, for instance, what products you might like on Amazon or what music you might enjoy on Spotify. According to a recent report in *JAMA Dermatology*, a machine-learning algorithm that distinguishes between images of benign and malignant moles has the potential to spot skin cancers missed by dermatologists.⁶ Early skin cancer diagnosis could particularly aid Black patients, who are less likely than White patients to develop melanoma but are more likely to die from it. However, the machine learning algorithms have been trained largely on examples drawn from White patients and are only now being designed in a way that would help control for potential bias.

In this article, I assess changes in the information and data ecosystem that should enable policymakers, researchers, and clinicians to harness this ubiquitous information architecture to identify health disparities, provide a method for evaluating them, and create effective interventions. I point out developments that allow this new technology to move forward at a very fast pace. In addition, I strike a cautious note about the potential negative consequences of the fourth industrial

revolution—consequences that may turn advances into a double-edged sword.

Advances in Data & Computing

Recent improvements in society’s ability to store and retrieve information, communicate rapidly via digital networks, and analyze data using increasingly powerful methods have fundamentally enlarged the country’s capacity to assess and intervene in health disparities. If policymakers, researchers, and clinicians take full advantage of this powerful combination of factors, they will be able to describe people’s health in multiple dimensions simultaneously and access information as needed. Combined with navigation systems at the personal, neighborhood, or community level, new analytic data capacities could identify, deal with, and remeasure health problems in a previously unimaginable time frame.

Newly available data sets contain immensely more information about individuals than is currently found in personal health records and other transactions captured by the health care system. Researchers and clinicians are now able to amass novel kinds of biological data, such as an individual’s genetic code. With the price of whole-genome sequencing dropping dramatically,⁷ scientists can envision a time when this information, consisting of more than 3 billion base pairs, will be routinely available as part of a person’s health record. The collection of such biomolecular data could lead to analyses that provide significant insight into the impact of innate biology on a person’s health and responses to the physical and social environments.

But purely genomic information is only a small part of the data that can increasingly be used to construct a biomolecular profile. A profile can also include information drawn from transcriptomics (the study of RNA molecules), metabolomics (the study of molecules involved

in metabolism), and analyses of the detailed workings of the immune system,^{8,9} in addition to integrative physiological information, such as heart rate and blood pressure, that can be measured with digital sensors. Although definitive evidence indicates that social determinants outweigh genetic influences on health risk at a population level, strong evidence also shows that biology has an impact on individual disparities in disease susceptibility and outcome. Further, as described below, when complex data become less expensive to collect, taking a sample or image, recording the information, and digitally storing it for later use becomes much more feasible. Thus, key outputs of a successful data-intensive approach to health disparities will include the delineation of biological mechanisms by which disparities lead to poor health outcomes as well as the development of interventions able to counteract those mechanisms—as was seen in AIR Louisville’s efforts to improve asthma control. Such work will also enable the planning of interventions that simultaneously deal with biological and social determinants of health—for example, an asthma intervention that involves both using medications and improving the home environment.

A fast-growing area of health measurement is called *digital phenotyping*,¹⁰ which characterizes people based on the way they interact with

cell phones, computers, and other personal devices. This information is deeply informative of the 99% of their lives spent outside of clinics and hospitals. Readily available and increasingly inexpensive sensors in cell phones and watches collect detailed information about individuals continuously over long intervals. Wearable sensors can measure activity levels, tremors, gait, and flexibility. Analyses of keyboard use and gait provide a deep measure of cognitive function, mood, and physical function. Use of cell phone apps and associated social media can provide detailed insights into social activity. Given the dominance of wealth, education, race, and location as mediators of health outcomes, the ability to directly measure behavior and social interactions will provide insights that could not be gained by asking patients questions during visits to a clinic or study site. For example, if your goal is to reduce cardiovascular disease in a population, it may be more important to insert green space and healthy food into neighborhoods than to increase the number of medical clinic visits.

The geospatial dimension is a particularly important factor in health disparities. It generally holds true that the most important predictors of health are zip code and income.^{11,12} Current technology allows health outcomes and determinants to be measured at a more granular level: household, street, neighborhood, county, and state levels. This type of measurement feeds into a potential understanding of social networks but also provides a substantial opportunity to make changes in the delivery system for both traditional medicine and social services and then feed information about those changes back to residents of affected neighborhoods and to medical clinicians and social service providers. As the speed of information acquisition, access, and analysis continues to increase, it will be possible to craft interventions at geographical (for example, neighborhood) or social (for example, workplace, school, and church) levels and measure outcomes to fine-tune the interventions (see Figure 1).

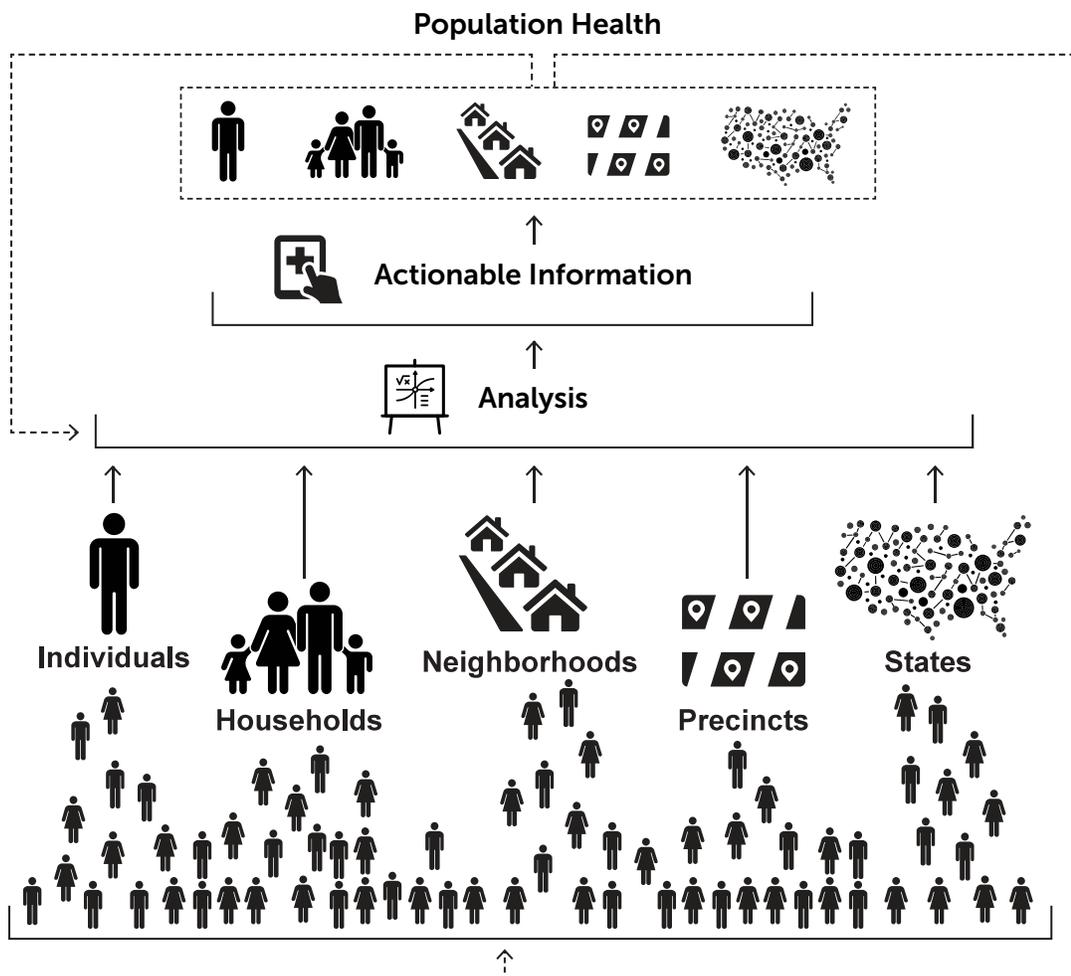
The dimension of time is also especially important in gathering individual data. In the past, clinicians and clinical researchers who

Key Priorities as Identified by Stakeholders in the Durham Health Innovations Project

- Increase health care coordination and eliminate barriers to services and resources.
- Integrate social, medical, and mental health services.
- Expand health-related services provided in group settings.
- Leverage information technology.
- Use social hubs such as places of worship, community centers, salons, and barbershops as sites for the distribution of clinical and social services and information.
- Increase local access to nurse practitioners, physician assistants, and certified nurse midwives.
- Use traditional marketing methods to influence health behavior.

Note. From "History," by Durham Health Innovations, n.d. (<https://sites.duke.edu/durham-healthinnovations/history>). Copyright 2015 by Duke Division of Community Health. Reprinted with permission.

Figure 1. Leveraging diverse sources of data to improve population health



Note. Information from across the spectrum of individual biology, clinical records, behavior, and social interactions can be combined with information about the fixed environment and the ambient environment (bottom). This information is summarized across individuals to produce population-based health measures that can be shared at the levels of the individual, the household, the neighborhood, the voting precinct, and the state. These lead to analysis and actionable information (middle sections). With appropriate analyses, evidence-based policies and interventions that are implemented (top) and assessed with repeated measures (dotted line on the left) can be devised to document the degree of effectiveness of the interventions. The information on more and less effective interventions can then be communicated to individuals and groups such as schools, churches, and neighborhood groups (dotted line on right). The partnership of these stakeholders is essential to the success of the effort.

wanted to evaluate changes over time were limited mostly to measurements made during periodic clinic visits. Digital technologies and the massive increase in the ability to manage data make it possible to use both passive and active data collection to accurately portray the impact of time. Passive measurements, such as the capture of heart rate and physical activity on smart watches and wearable fitness monitors, can be obtained almost continuously. Cell phones can also be used to collect frequent passive impressions from the research participant or patient. As discussed above, this radical reduction in data latency is also pertinent for

interventions at the group level, as evidenced by the use of social media to deal with environmental catastrophes.

The study of sleep offers an example of the potential importance of collecting data that cover temporal, spatial, and behavioral dimensions. In the past, this kind of research has relied on either patient recall or intensive study in sleep units, which are artificial environments with little resemblance to the home environment. Now, with passive sensors on wearable technology, the quantity and quality of sleep and attributes of the home environment can be

measured without disrupting normal patterns and without great expense to the patient. It is likely that many of the medical, social, and environmental factors that affect sleep are also associated with differences in longevity. Enabled by a system of integrated information such as the one described above, a clinician could intervene at the individual level by prescribing weight loss procedures (to reduce sleep apnea) or sleep medication. Medical institutions, health planners, and legislators could make changes at the group level through social networks or at the geographic and environmental levels by, for instance, reducing noxious sounds by changing traffic patterns.

650

Health care systems that account for 90% of all hospital discharges



There are approximately 3 billion base pairs in the human genome.

60%

Health Disparities

Variation in countywide life expectancy due to socioeconomic and race/ethnicity factors

Harnessing Big Data

The consolidation of clinical care systems, in which big health systems absorb smaller ones, provides opportunities to measure and change health disparities. As of several years ago, the Agency for Healthcare Research and Quality estimated that approximately 650 health systems account for over 90% of hospital discharges and are increasingly integrating hospital and outpatient care, including assisted living, nursing homes, and hospice.¹³ These systems are developing sophisticated data “lakes” and warehouses to aggregate information that is curated so that it can show how health care can be delivered in a financially sustainable manner. (The word *curated* deliberately evokes the way historical archives work. Curated data include critical accompanying information about a data point’s context—where it was gathered and how it was gathered, stored, transmitted, and transformed, and by whom.)

Because these health systems need standardized information to conduct business, data in several dimensions need to become integrated in a variety of ways. First, the various entities within each system must use common standards and definitions for the system to function efficiently and for patients to benefit from the sharing of information across practices. Second, health care delivery system data should be integrated with information from other dimensions that are pertinent to health—for example, social and environmental factors such as the

presence of green spaces, the availability of physicians’ offices, and traffic patterns.² Third, multiple projects have shown that these data can be merged to produce data sets that can be easily shared across systems, with the goal of improving outcomes.

The National Patient-Centered Clinical Research Network (PCORnet; <http://pcornet.org>) illustrates the magnitude of data integration that is possible. This project, funded by the Patient-Centered Outcomes Research Institute, has developed a systematic approach to curating data across multiple health systems, including both clinical care data derived from electronic health records and insurance claims data, for over 100 million Americans. PCORnet is now evolving into the People-Centered Research Foundation (<http://pcrfoundation.org/>), a not-for-profit organization dedicated to conducting pragmatic randomized trials (research performed not at an artificial research center but at a real patient point of contact with an eye to informing decisionmakers of the comparative balance of benefits, burdens, and risks of a biomedical or behavioral health intervention) and observational studies (nonrandomized studies that allow for historical comparisons).

One interesting component of the People-Centered Research Foundation is the ADVANCE Collaborative, a network of federally qualified community health centers.¹⁴ This organization includes OCHIN, Health Choice Network, Fenway Health, Kaiser Permanente Center for Health Research, Legacy Health, Care-Oregon, and Oregon Health and Science University. Intended as a learning laboratory for policymakers to better understand patients who use safety-net services, the system has digital records on more than 3 million patients, including large numbers of homeless, uninsured, underinsured, and undocumented people, as well as members of other underrepresented populations. This type of network, if coordinated with more traditional integrated health care systems, could provide a mechanism not only for evaluating disparities but also for designing and testing interventions, such as drug counseling, at the system level.

Opportunities for Data to Characterize Disparities

Several organizations are already providing analyses of data that depict health disparities in intuitively understandable displays. Perhaps the most far-reaching of these reports is put out by collaborators working on the Global Burden of Disease Study.¹⁵ Another recent series of reports done mainly at Harvard University has clearly demonstrated that variation in longevity and disease burden in the United States is a function of geographic location at the county level.¹⁶ The striking impact of residence in rural counties is highlighted by the visual depiction of both current longevity and trends in longevity over time. One recent report in this series determined, for instance, that 60% of the variation in countywide life expectancy is explained by socioeconomic and race and ethnicity factors and that rural counties fare worst on such measures as mortality rates, suicides, drug overdose deaths, rates of teenage pregnancy, and fetal and maternal mortality.¹² Detailed analyses of these data have demonstrated that, as expected, wealth, education, race, sex, and location are key factors in longevity,¹² mediated in common chronic diseases by factors such as blood pressure, low physical activity, tobacco use, obesity, depression, and diabetes mellitus.

Many organizations routinely produce comparative reports of health status. Within the United States, significant efforts are aimed at curating actionable data at the level of the city, county, or state. Some of the most potent information comes from the evaluation of boroughs in New York¹⁷ and from state-level reporting by the Robert Wood Johnson Foundation.¹⁸ Research in Durham County, North Carolina, demonstrated the power of this information when it was applied at the level of individual households and neighborhoods to, among other goals, reduce exposure to lead poisoning.²

New Ways to Use Data to Reduce Disparities

The same information infrastructure used for measurement could also be used for implementing interventions (See *Key Actions Needed to Collect and Use Actionable Data to Reduce*

“it should be possible to provide useful information in a way that is tailored to the specific health needs of an individual or family”

Health Disparities). At the level of a community or health system, interactions between traditional health systems and social services tend to be inefficient. Improved labeling of government, private, and volunteer services and coordination of these services with clinics, schools, and businesses could lead to a much more directly effective intervention system.

In addition, the ubiquity of cell phones and steep reductions in the cost of sensors make it possible for clinicians to communicate directly with individuals and groups at any interval that is desired. Additionally, almost all people use search engines to seek information on a routine basis. Search engine results are tailored by machine-learning algorithms to an individual's pattern of communication. As the curation and organization of information continues to improve, it should be possible to provide useful information in a way that is tailored to the specific health needs of an individual or family. For instance, these approaches can be used to fine-tune search results, much as consumer goods are currently surfaced in a manner consistent with the preferences of the consumer. For example, when someone searches on the term *stage 1 breast cancer*, it is technically possible for the high-ranked results to be tailored to the medical literacy of the individual as well as authoritative, relevant, and trustworthy. For the most common health searches, Google is currently providing “knowledge panels” that are vetted by medical experts.

Consider the ongoing epidemic of asthma. Asthma is often exacerbated by environmental triggers, both within and outside the home. Further, research has demonstrated that

disparities in asthma incidence and access to care are functions of wealth, education, physical location, and race.¹⁹ Inexpensive sensor technology and ubiquitous data networks would enable clinicians to monitor environmental quality at the household and neighborhood levels, which would make it possible for them to deploy precise interventions to reduce stimuli that exacerbate asthma. The previously mentioned report from the study in Louisville³ points out that this sort of intervention, which focuses on cleaning up the home and neighborhood environments, would potentially be much

more powerful than medication in preventing asthma exacerbations.

Similarly, obesity and diabetes contribute to an enormous amount of death and disability, and the geographic and social profiles of relevant health disparities are clear.²⁰ Although special medical clinics, surgical and medical interventions, and wide dissemination of accurate and useful information that reaches the people who need it are all possible solutions, there is ample reason to believe that constant exposure to advertising for food, long distances to grocery stores that sell healthful food, and cultural and environmental influences on physical activity limit the success of medical interventions for people with lower incomes or other socio-economic disadvantages. A wealthy, highly educated person who can afford a personal trainer or gym membership and is not caring for family members is more likely to be able to engage in a healthy lifestyle. Health systems, advocacy groups, community leaders, and individuals at all levels of government can engage with people more productively within their personal digital environments, helping them to use geospatial information to locate healthful food resources and enabling and encouraging them to integrate physical activity into their routines in a more economically feasible manner that intrudes less into other aspects of life.

Key Actions Needed to Collect & Use Actionable Data to Reduce Health Disparities

Engage people and communities as partners

- Requires face-to-face time and use of social media
- Transparency is critical at all steps
- Issues of privacy and confidentiality require considerable work

Collect diverse sources of data

- Biological, clinical, behavioral, social, and environmental data are needed
- Full data use will require solutions to engagement, partnership, and privacy/confidentiality issues

Curate and organize data

- Curation and organization are currently the most underinvested area of data science, requiring significant investment
- Requires conscious investment at the institutional level by health systems and government entities

Analyze

- Methods involving geospatial orientation and hierarchical analysis from the level of the individual to population will be informative
- These data are big
- Speed of access to data and fluidity of data are critical factors in making data actionable
- Identifiable data will be most actionable but also riskiest from a privacy/confidentiality perspective

Use outputs of analyses to formulate policies

- Requires collaboration across health systems, neighborhoods, and policymakers at local, state, and national levels
- Participants in the effort need education on quantitative and community engagement methods

Implement policies

Measure again and adjust on the basis of outcomes

Importance of Community Engagement & the Development of Shared Approaches

One area that researchers need to study more thoroughly is how to best transmit new information directly to those who are affected by it as well as to those who can implement interventions and policies to improve outcomes. Although research that engages communities²¹ continues to advance, and many communities are involved in direct interventions, policymakers still lack clarity on which methods are likely to be most effective at linking personal health data with social and environmental information in ways that yield measurable improvement in outcomes. A promising approach has been developed by the Abdul Latif Jameel Poverty

Action Lab, which is using observational, experimental, and quasi-experimental methods to understand which social policies lead to improved outcomes.²² One such study showed that charging fees for preventive medicine tools in low-income countries drastically reduces their usage.²³

Over a decade ago, Durham County, North Carolina, and Duke University initiated an ambitious program that showed how to create successful communication between government, health institutions, and the community. Called Durham Health Innovations (DHI),²⁴ the program existed in the context of a history that included decades of both outstanding collaboration between the university and the community and well-documented divisions and disparities.²⁵ For the project, teams of volunteers were organized so that they had equal representation from the community and the university (including its health system and academic medical center). Each team was focused on addressing a particular health issue of significant concern to the community. Teams were then supported with the data assets of the academic health system, the Durham County public health department, and other entities and asked to devise an approach to health care and community intervention that would improve health outcomes relevant to the issue they had chosen. The teams winnowed their issues down to 10 major problems affecting the community: adolescent health; asthma and chronic obstructive pulmonary disease; cancer prevention and early detection; cardiovascular disease; diabetes; HIV, sexually transmitted diseases, and hepatitis; maternal health; obesity; pain management; and healthy aging in place.²⁴

Remarkably, after a series of meetings and discussions aided by intensive data analysis, the proposed approaches to these seemingly different problems all converged on a common set of interventions that could improve the health and health care delivery in Durham regardless of disease or therapeutic area (see *Key Priorities as Identified by Stakeholders in the Durham Health Innovations Project*). The project was ahead of its time and in many ways anticipated the wayfinding approach mentioned

“it is critical that data holders discuss their intentions with the data providers”

in the introduction. It pulled in diverse sources of data to provide a holistic understanding of health needs, with geographic information integrated to guide intervention. But because health data do not enjoy the kind of fluidity and latency advantages leveraged by wayfinding apps, DHI was limited in what it could accomplish. Some of the critical roadblocks to implementing fourth industrial revolution–style solutions can be overcome once project directors can access the right data, in the right way, at the right time.

Steps to Limit the Risks of Data Sharing

Sharing data between patients, physicians, and institutions requires a degree of trust. And clearly, for all the good data sharing can do, a markedly enhanced system of measurement, assessment, and intervention could be used for negative as well as positive purposes. For instance, Facebook’s brokerage of personal data during the 2016 U.S. presidential election showed that social media data could be put to nefarious uses.²⁶ Theoretically, health care information, perhaps hacked from a medical center’s patient portal, could be leveraged to, say, target underinsured cancer patients with an ineffective but expensive “cure.” False news is particularly dangerous. Empirical evidence is accruing that false statements, many of which are intended to stoke differences among people, reach more people faster and persist longer than truthful statements do because their novelty gives them “legs.”²⁷ These issues are not new, but they have emerged as critical considerations in deciding how to use information to improve the outcomes of those who suffer from health disparities.

To increase transparency, engagement, and trust by patients, clinicians, and institutions, it is critical that data holders discuss their intentions with the data providers. In a body of work

on the moral obligations of health systems that are continuously evaluating accruing data to improve health care and health outcomes—that is, *learning health systems*—ethicists Nancy E. Kass and Ruth R. Faden explore a crucial fundamental concept that bears on these challenges: the reciprocal obligation of those who use data to those who provide the data.²⁸ Although these concepts are reasonable, they have not yet been fully implemented. It is interesting that Kass and Faden’s scheme considers reducing disparities to be an essential element of learning health systems.

Given the complexity of interactions needed to successfully implement interventions that reduce health disparities, it will be important for multiple societal entities to involve themselves in developing the cultural and legal expectations that will enable big data to be used effectively for desirable purposes. For example, the current HIPAA laws deal with health care data in a manner that many consider to be overly restrictive, whereas the data from “the other 99%” of life, which have a much larger impact on health outcomes, are governed by much less restrictive rules. Perhaps it would be better to have a single data standard that is pertinent to health outcomes. For this expectation-setting effort to succeed, societal entities that are not motivated by profit and are capable of convening diverse interests should help to devise standards. Universities are in a special position to engage with communities and offer the benefits of faculty knowledge and skills in medicine, law, technology, and ethics as they work with community groups and individuals to devise policies that will achieve the desired results.

A Possible Future

The amount of information now available about individuals and their health—constantly generated and recorded by new, interconnected devices—is multiple log orders more voluminous and complex than the data from patient interactions currently available to clinicians, health care administrators, policymakers, and clinical researchers, and the cost of managing the data is rapidly declining. It is technologically possible to observe, describe, and analyze health disparities at the levels of individuals, households, streets, neighborhoods, cities, and counties. Information-based profiles at each of these levels could be further segregated by biological, medical, behavioral, social, or environmental characteristics. The relevant information that could guide and measure interventions to improve health and reduce health disparities could be displayed in any time interval desired. Interventions could be planned at any of these levels and the results measured in ways that would reveal cause and effect and suggest useful interventions. These same methods, however, if applied for selfish or ignoble purposes, could be leveraged to increase health disparities. Nevertheless, here at the leading edge of the fourth industrial revolution, new methods of data curation and analysis can provide the foundation for a dramatically improved approach to health disparities for both individuals and populations.

author affiliation

Califf: Duke University School of Medicine, Stanford University School of Medicine, and Verily Life Sciences. Corresponding author’s e-mail: robert.califf@duke.edu.

references

- Schwab, K. (2016). The fourth industrial revolution: What it means, how to respond. Retrieved from <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>
- Miranda, M. L., Ferranti, J., Strauss, B., Neelon, B., & Califf, R. M. (2013). Geographic health information systems: A platform to support the "triple aim." *Health Affairs*, *32*, 1608–1615. <https://doi.org/10.1377/hlthaff.2012.1199>
- Barrett, M., Combs, V., Su, J. G., Henderson, K., Tuffli, M., & the AIR Louisville Collaborative. (2018). AIR Louisville: Addressing asthma with technology, crowdsourcing, cross-sector collaboration, and policy. *Health Affairs*, *37*, 525–534. <https://doi.org/10.1377/hlthaff.2017.1315>
- World Bank. (2016). *World development report 2016: Digital dividends*. <https://doi.org/10.1596/978-1-4648-0671-1>
- BBC News. (2018, August 24). Russia trolls "spreading vaccination misinformation" to spread discord. Retrieved from <http://www.bbc.com/news/world-us-canada-45294192>
- Adamson, A. S., & Smith, A. (2018). Machine learning and health care disparities in dermatology. *JAMA Dermatology*. Advance online publication. <https://doi.org/10.1001/jamadermatol.2018.2348>
- National Human Genome Research Institute. (2016). The cost of sequencing a human genome. Retrieved from <https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/>
- Institute of Medicine. (2014). *Capturing social and behavioral domains in electronic health records: Phase 1*. Retrieved from <http://www.nationalacademies.org/hmd/Reports/2014/Capturing-Social-and-Behavioral-Domains-in-Electronic-Health-Records-Phase-1.aspx>
- Institute of Medicine. (2014). *Capturing social and behavioral domains in electronic health records: Phase 2*. Retrieved from <http://www.nationalacademies.org/hmd/Reports/2014/EHRdomains2.aspx>
- Insel, T. R. (2017). Digital phenotyping: Technology for a new science of behavior. *JAMA*, *318*, 1215–1216.
- Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., . . . Cutler, D. (2016). The association between income and life expectancy in the United States, 2001–2014. *JAMA*, *315*, 1750–1766. <https://doi.org/10.1001/jama.2016.4226>
- Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., Mackenbach, J. P., van Lenthe, F. J., . . . Murray, C. J. L. (2017). Inequalities in life expectancy among US counties, 1980 to 2014: Temporal trends and key drivers. *JAMA Internal Medicine*, *177*, 1003–1011. <https://doi.org/10.1001/jamainternmed.2017.0918>
- Agency for Healthcare Research and Quality. (2016). *Compendium of U.S. health systems, 2016*. Retrieved from <https://www.ahrq.gov/chsp/compendium/index.html#data>
- Patient-Centered Outcomes Research Institute. (2017). Accelerating Data Value Across a National Community Health Center Network (ADVANCE). Retrieved from <http://pcorner.org/clinical-data-research-networks/cdrn10-oregon-community-health-information-network-ochin/>
- GBD 2016 Causes of Death Collaborators. (2017). Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet*, *390*, 1151–1210.
- Murray, C. J. L., Kulkarni, S. C., Michaud, C., Tomijima, N., Bulzacchelli, M. T., landiorio, T. J., & Ezzati, M. (2006). Eight Americas: Investigating mortality disparities across races, counties, and race-counties in the United States. *PLoS Medicine*, *3*, 1513–1524. <https://doi.org/10.1371/journal.pmed.0030260>
- New York City Department of Health and Mental Hygiene. (n.d.). New York City community health profiles. Retrieved from <https://www1.nyc.gov/site/doh/data/data-publications/profiles.page>
- Robert Wood Johnson Foundation. (n.d.). County health rankings & roadmaps. Retrieved from <https://www.rwjf.org/en/how-we-work/grants-explorer/featured-programs/county-health-ranking-roadmap.html>
- Volerman, A., Chin, M. H., & Press, V. G. (2017). Solutions for asthma disparities. *Pediatrics*, *139*, 1–9. <https://doi.org/10.1542/peds.2016-2546>
- Centers for Disease Control and Prevention. (2014). *National diabetes statistics report: Estimates of diabetes and its burden in the United States, 2014*. Retrieved April 2, 2015, from <https://www.cdc.gov/diabetes/pdfs/data/2014-report-estimates-of-diabetes-and-its-burden-in-the-united-states.pdf>
- Michener, L., Cook, J., Ahmed, S. M., Yonas, M. A., Coyne-Beasley, T., & Aguilar-Gaxiola, S. (2012). Aligning the goals of community-engaged research: Why and how academic health centers can successfully engage with communities to improve health. *Academic Medicine*, *87*, 285–291.
- Abdul Latif Jameel Poverty Action Lab. (n.d.). About us. Retrieved from <https://www.povertyactionlab.org/about-j-pal>
- Abdul Latif Jameel Poverty Action Lab. (2018). The impact of price on take-up and use of preventive health products. Retrieved from <https://www.povertyactionlab.org/policy-insight/impact-price-take-and-use-preventive-health-products>
- Durham Health Innovations. (n.d.). History. Retrieved from <https://sites.duke.edu/durhamhealthinnovations/history>
- Baker, J. P. (2017, September 19). A tale of two cities? *The history of Duke Health and Durham's health* [Video file]. Retrieved from <https://dukemed-source.mediasite.com/Mediasite/Play/6b89b77ef0fa465daa848b6e139653ce1d>
- Solon, O. (2018, April 4). Facebook says Cambridge Analytica may have gained 37M more users' data. *The Guardian*. Retrieved from <https://www.theguardian.com/technology/2018/apr/04/facebook-cambridge-analytica-user-data-latest-more-than-thought>
- Vosoughi, S., Roy, D., & Aral, S. (2018, March 9). The spread of true and false news online. *Science*, *359*, 1146–1151.
- Faden, R. R., Kass, N. E., Goodman, S. N., Pronovost, P., Tunis, S., & Beauchamp, T. L. (2013). An ethics framework for a learning health care system: A departure from traditional research ethics and clinical ethics. *The Hastings Center Report*, *43*(Suppl. 1), S16–S27. <https://doi.org/10.1002/hast.134>