

WHY BIG DATA WON'T CURE US

Gina Neff

*Department of Communication
University of Washington, Seattle, Washington*



THE BIGGEST CHALLENGE for the use of “big data” in health care is social, not technical. Data-intensive approaches to medicine based on predictive modeling hold enormous potential for solving some of the biggest and most intractable problems of health care. The challenge now is figuring out how people, both patients and providers, will actually use data in practice.

To understand how data-intensive solutions could have an impact on health care, our research team talked to frontline providers in impoverished and rural areas, technology enthusiasts in mobile health and health IT startups, clinicians and researchers in major research hospitals, Quantified Self members at data-driven meetup presentations of massive amounts of tracking data, and attendees at the growing number of conferences for health technology and innovation up and down both coasts. I found the buzz as feverishly loud around health information innovation as it was during my research on the first dot-com boom.

One of our findings from this research seems at first blush so obvious that it is hard to believe it has been overlooked in the design and implementation of health-care innovation technologies. Namely, people imagine *data* in very different ways. Understanding this key fact about data helps us understand why so-called “big data” solutions to health care are so difficult to implement in practice. Doctors, patients, and health-care entrepreneurs all *value* data in very different ways. One physician simply said, “I don’t need more data; I need more resources.”* Saying this in Silicon Valley or at TedMed would be tantamount to heresy. Ditto for those of us who work in research and spend our careers collecting, massaging, man-

aging, analyzing, and interpreting data. From the doctor’s perspective, though, data require (and do not save) extra interpretive, clerical, and managerial labor. This perspective on “data,” at least with regard to current clinical practice, is that data use up more resources than the benefits they provide. In other words, most doctors think data innovation means more work for them, not less, and takes away time from what they see as their key priorities in providing quality care.

In another setting, we observed nurse-practitioner case managers in a Medicare demonstration project working with a simple algorithm parsing patient-entered health data. Combined with case management, these data provided a look into the daily health of chronically ill elderly patients and a pathway for the care when it was needed. The data in that project were tightly tied to medical expertise within an existing clinic where a trusted person could initiate a chain of care responses. Although widely recognized as a clinical success, Medicare pulled the plug on the project for financial reasons—expertise is expensive.

These two reactions to data-intensive pilot projects highlight the dilemmas of data-analytic approaches to health care. Businesses are in the thrall of the possibilities of ever-increasing predictive analysis on expanding troves of generated data. While the business and technology sectors see data as valuable, doctors often see data as costs, risks, and liabilities. And for many in health care, data are not seen as a source of value, but of additional work. Without the work needed to make data valuable and useful in particular settings in particular contexts in health care, big data will never solve

*Author’s interview data. See Neff & Fiore-Silfvast n.d. for more details.

problems. To turn a technology truism on its head, data in health care will never be free.

And yet, the ways in which health technology innovators have talked about the power of data neglects key aspects of the *social interoperability* or integration of data into health solutions. How will such data be integrated into care providers' work practices; through the complex routines of clinics and hospitals; and into existing legal, social, political, and economic frameworks? These questions are enormous. Until we solve these questions of social interoperability, the risks presented by "big data" in health will outweigh the benefits to any particular individual, regardless of whether we're talking about terabyte-scale analytics or the "small-data" of $n=1$ individuals. What follows is an outline of how to tackle these questions based upon what our team has seen throughout our research.

Why Think About Big Data in Health?

Two types of data are generating excitement for application in the health-care field. The first type is big data, or analytics of multiple types of data across a population, potentially from multiple sources, of structured and unstructured nature (i.e., readily merged into traditional database structures or not) and of heterogeneous kinds of objects (text, numbers, images, documents, locations). The hope for such data is simply stated by Ginger.io—one of the many rapidly emerging startups in this field because of what they might reveal across many people—in their tagline "Big data, better health."¹ Such data are neither new nor novel, but more are being generated. This is not so much a new kind of data but huge amounts of it. Assembled and analyzed, it could all be a potential valuable resource. Or it could be what privacy guru Bruce Schneier calls the "pollution" of the information age, a byproduct produced by virtually every technological process,² something that is more costly to manage than its value.

The second type is what Cornell NYC Tech professor Deborah Estrin has called "small data," the output of a whole host of pervasive tracking processes about any one individual user.³ On the avant-garde of using this type of data are those involved in the "quantified self," or QS movement, who enthusiastically measure and track a variety of aspects of their everyday lives. As any reader of Benjamin Franklin knows, personal tracking is not new, but it is made newly relevant and accessible with the possibilities presented by ubiquitous and pervasive computing of smartphones, digital activity

trackers, Wi-Fi-enabled scales, and other such devices. Melanie Swan's recent review in this publication addresses the potential of such small data approaches, technologies, and practices for creating the QS founders term an "exoself," a digital representation modeling the body, the self, and the behavior.⁴

Such "big" and "little" data both hold exciting possibilities for the discovery of patterns. Patterns in this data can be inspiring, wondrous, curious, surprising, and yet frustratingly tough to interpret. Big data approaches to health-care research promise the possibility of larger study populations than ever thought possible. Advances in computing and communication mean more, and different types of data can be linked and analyzed across more people in novel ways. Recent initiatives such as the Health Data Exploration Project, based at the University of California at San Diego (UCSD) and backed by the Robert Wood Johnson Foundation's

Pioneer Portfolio, are trying to figure out practically and ethically how to scale up and aggregate personal health data across many individuals for larger-scale research.⁵ The potential for the discovery of new, previously unseen connections makes data-intensive health an exciting research frontier, and the possibility of contributing to discovery and the public good may be a big motivator for people in allowing their data to be used by researchers. But data discoveries do not necessarily benefit the same people generating, producing, and sharing the most minute and intimate details of their lives. Data-driven discovery in health care may produce a public benefit that comes at a cost or risk to the privacy of the individuals who make such discovery possible.

Applying cutting-edge research from any domain to routine clinical care is challenging. As always, just because there is exciting research that presents a lot of possibilities, it doesn't mean that a very traditional set of social institutions will change. We can talk about how big data may disrupt health care in the future, but what we have found through our research is that the established ways of practicing and organizing health care are deeply entrenched.

Big Data Solutions Are Not Yet Connected to Care

To use a medical metaphor, "bench science," or data-intensive health research, is currently further advanced than "translational science," or the clinical practice of data-

**"I FOUND THE BUZZ AS
FEVERISHLY LOUD AROUND
HEALTH INFORMATION
INNOVATION AS IT WAS DURING
MY RESEARCH ON THE FIRST
DOT-COM BOOM."**

intensive medicine. We need now to connect data-derived insights to clinical care and translational medical expertise. Whether data are gathered across a population or for a consumer's own personal use, there exist few mechanisms for using these types of data as resource for the diagnosis and care of individuals. Making these types of data socially interoperable means understanding the differences in how people generate, use, and even talk about data.

Big data has a rhetoric problem. When people talk about data-driven health innovation they often neglect the power of framing information as “data.” They also assume that everyone thinks about health data the same way they do. Regardless of how it is generated, digital information only becomes data when it is created as such. Calling traces of digital behavior or personal histories *data* masks bigger questions: data for whom and what purposes; data when and data why? Information useful for the online marketer is not necessarily useful for the patient or the clinician or the researcher. Data are meaningful because of how someone collects, interprets, and forms arguments with it. Data are not neutral. This is why Lisa Gitelman calls raw data an “oxymoron,” a contradiction in terms that hides the reality of the work involved in creating data.⁶ Data, I argue in an article with Brittany Fiore-Silfvast, are so important precisely because people make (or imagine) data function across multiple social worlds.⁷ Data are not inherently important or interesting, rather, by definition, data are used to make arguments relative. Put simply, data is only data in the eye of the stakeholder.

Take patients' own mobile health and wellness data. Patients feel, in part, that such data are significant because they reflect their stories and provide opportunities to connect and converse with care providers and others. Consumer-directed (as opposed to medically regulated) mobile apps offer new ways of encouraging and supporting the behavioral changes that improve health, whether or not medical expertise is brought to the data. A primary care doctor isn't likely to be interested in routine pedometer readings for most of her patients but is very interested in encouraging sedentary patients to be more active. Here, data are not as useful for diagnosis or clinical decision making (although we can imagine several scenarios in which they could be) as they are for self-reflection and individual change. Yet, when clinicians talk about data, they tend to prioritize what data can do clinically for diagnosis, treatment, or decision making. Data from mobile health and wellness applications may have little utility for clinical decisions. Individuals' mobile health data may be problematic for their health-care providers because these data bring issues of

reliability, liability, and cost to the clinician for clerical and diagnostic time, with little promise—at least for now—of improving clinical outcomes. This is one example of how different stakeholders have different expectations for what they call data.

Even as people's use of digital media for getting and sharing health information increases, such data are not yet routine parts of conversations between patients and their providers. Patients are using digital media for social support with family, friends, and people with similar conditions, but cannot use these tools to communicate with their health-care providers. Translating what counts as data across the social worlds of patients and health-care providers is the first step. Sharing data across the social worlds of patients and providers is one of the significant obstacles in big data health.

Researchers have these data translation problems, too. A biostatistician highlighted this difference between the different data cultures of clinicians and researchers: “Physicians are typing away madly. All that information is actually very

rich.”⁸ Where a biostatistician may have seen value in physician's notes for a while, the rest of us are only now coming to see that they might be data that can be mined for value. On the other hand, contexts, not just numbers, matter to health-care providers. This was evident when we watched how clinical care case managers talked about and made allowances for algorithmically

parsed data about their patients. They explained variation within and across the numbers of each of their cases from first-hand experience and interpersonal interactions, in effect doing data interpretation on the fly. As one clinical information systems researcher put this adjudication between data and context, “A computer usually looks at one small aspect of the patient's problem but doesn't get the context. An expert doctor can understand the huge picture of what's going on with a patient.”⁹ People in the different social worlds of health care—such as lab analysts, startup entrepreneurs, clinical health-care providers, patient, online consumers, and insurers—all think of data differently and do (or hope to do) varying work with that data.

Currently, we don't have very good bridges for data to cross these social worlds of health care. The routines and practices of the clinical care for patients are not connected, for the most part, to data analytics. The guidelines being issued by the Office of the National Coordinator for Health IT go a long way in providing the kind of leadership necessary for translational data science.¹⁰ For example, the first of these guidelines maps how hospitals and other stakeholders can use

**“DATA ARE MEANINGFUL
BECAUSE OF HOW SOMEONE
COLLECTS, INTERPRETS, AND
FORMS ARGUMENTS WITH IT.
DATA ARE NOT NEUTRAL.”**

analytics of their patients' electronic health records to reduce hospital readmission rates. These data-intensive strategies may work better for larger organizations with more resources than in smaller clinics or for individual doctors. Care providers do not have the time, expertise, or resources to utilize predictive analytics or quantified self-inspired metrics for patient care, and translating research into clinical care takes time and work. There is little time to go over data from medical devices in the course of a clinical visit, much less from the plethora of lifestyle devices that are being marketed for self-tracking.

In the course of our research we heard several young, tech-savvy diabetic patients present sophisticated analysis of the data generated from their continuous glucose monitoring sensors only to report that it was difficult to talk with their doctors about the data. Experienced designer and self-quantifier Katie McCurdy summed up these difficulties in her own attempt to bring data to her doctors in a presentation on the quantified patient, "I want to work with a doctor who believes me."¹¹ When chronically ill but engaged patients have difficulty getting time to discuss their data with their doctors, it does not bode well for people who want to jointly interpret other kinds of potentially rich data. Many of the most exciting new tools to date have been designed without considering how doctors and patients communicate. Before we can talk about the integration and analysis of multiple data sources in electronic health records, we must figure out how these data can be used by patients with their doctors for their joint decision making in practice.

Is Knowledge the Answer?

Much of the "small" or individual-scale data discussed in data-intensive health solutions is consumer oriented. Such data is useful in part because it's connecting people to others, either through comparison of their data or through social support for behavior change. These moves change the nature of what the data do, and as a result these data may not fit the mental models of how people imagine their data working for them. In research being done by Heather Patterson and Helen Nissenbaum of New York University, the privacy expectations that people have for such data, what they term "contextual privacy," shifts from lifestyle uses of wellness data and health-care uses of that data. In the context of lifestyle tools, privacy decisions are made differently than in a medical context, even when the underlying data are the same. Para-

doxically, people may be less guarded about sharing information with the for-profit companies that make their fitness apps than they are with HIPAA-bound health-care providers. This is in part because quantifiers need good information from their tools and devices for them to be useful in guiding their health and wellness choices and behaviors.¹²

Behavior modification models that lack in either nuance or sophistication may explain why the usage of mobile health apps drops off dramatically after downloading. The Pew Internet and American Life report on Mobile Health 2012 found 19% of smartphone owners and 11% of all mobile phone owners have a health application on their phone, but adoption rates of these applications remain flat. Still, nearly half of the surveyed adults reported that they tracked their health, including those who track on paper and "in their heads." Susannah Fox, one of the report's authors, likens the phenomena of dataless tracking to "skinny jeans" kept in the closet that a woman uses in lieu of a scale to gauge her weight.¹³ Such different relationships to numbers and knowledge must be considered in the design of health innovations targeted toward tracking and behavior change outside the motivated, tech-savvy quantifiers.

Similarly, people may be willing to share their data to benefit their own health and wellness practices or the public good of scientific discovery but may be more reluctant to do so to benefit commercial interests. And yet, one person's data may only be valuable in relation to that of others.* The usefulness of data across social worlds is relative and shapes the values that inform people's privacy choices. Health and wellness data design involves choices that have enormous implications for social justice, power, and autonomy, as well as for control over both the risks and benefits of the data.

The question remains: Who then *can* use the data?

Big Data, Justice, and Power

There are persistent myths about big data. According to Kate Crawford of Microsoft Research, it is dangerous to think that big data, including that large-scale data can be made anonymous, are inherently objective and include tacit or explicit consent or an opt-out function. Research by computer scientists continue to show that even anonymized data can be reidentified and attributed to specific individuals.¹⁴ A pioneering researcher in this field, Latanya Sweeney, was part of

"HEALTH AND WELLNESS DATA DESIGN INVOLVES CHOICES THAT HAVE ENORMOUS IMPLICATIONS FOR SOCIAL JUSTICE, POWER, AND AUTONOMY, AS WELL AS FOR CONTROL OVER BOTH THE RISKS AND BENEFITS OF THE DATA."

*For example, see the work being done at wethedata.org

a team that was able to link names and contact information to the publically available profiles in the Personal Genome Project through exploiting known weaknesses in large-scale demographic datasets.¹⁵ By mining public records for the seemingly less innocuous information of birth date, gender, and zip code, they correctly identified (84–97% of the time) anonymous profiles that contained demographic along with genetic and medical information.¹⁶ If data from just a few pieces of less-protected demographic information can re-identify someone, imagine what adding genetic information or disease conditions could mean for privacy risks in large-scale shared and pooled data.

Much of what is called big data in the commercial realm relies on forms of consent that make it difficult for people to understand the true nature of the risks to their privacy and difficult or impossible for them to opt out. Within the health-care sector there is a powerful drive to use patients' protected health data for the analytics on the business of health care, such as to improve hospital efficiency and health costs savings. There are real privacy risks to patients posed by these processes, while privileging benefits to stakeholders other than the patient, and no options being offered to opt out of such analytic uses of their information. Such decisions in the design of data solutions do more than fail to put the patients' interests first. They fundamentally shift the relationship of power and control between patients, their health-care providers, and the insurers around the questions of data. If the conversation in health technology innovation does not address the questions of data for whom, when, and why, then it will be a failure of social justice and an abuse of the trust that people have placed in the institutions of health care.

Conclusion: Real Data for Real Practices and Real Contexts

In the end, these are solvable problems for exciting times. Policy makers, advocates, and technology designers alike must remember that the solutions for the problems of health information innovation are as much social as they are technical. From that perspective, below are five elements that must be a part of any push toward big data health care.

1. Real conversations on data privacy: Several health IT entrepreneurs are calling for reforms to the Health Insurance Portability and Accountability Act (HIPAA) that don't thwart innovation, and the thicket of policies

discourages many entrepreneurs from working within the FDA regulatory system. There is simply too much at stake to risk privacy and security to have such a wide gulf between unregulated wellness and regulated health applications and information solutions. Likewise, the potential benefits of data-intensive approaches to health need not be derailed by solvable security concerns. As researchers learn more about the relationships of online data and behavior to health, we need to consider how we label, frame, and store many more types of data as patient data or not, and this is a conversation that is even bigger than the health care sector. Let's build apps that protect privacy, not water down privacy rules for the sake of more apps. Let's stop assuming anonymous data linked to demographic variables is actually anonymous and start protecting the data for what they are. And finally, let's begin a conversation around the rights of digital citizenship that restores power, transparency,

and control to people in a broader set of data interactions.

2. Design that matters for clinical care: Rather than recite empty rhetoric about the disruption of health care, designers must begin to include data's relationship to clinical care at the core of their design considerations. Solutions will not be found through simply aspiring to integrate electronic health records. Taking into consideration how data practices in the clinic benefit patients is an issue

of equity, justice, and social power. Designing for the challenges that health-care providers face will mean creating data solutions that dovetail and enhance existing medical knowledge and practices, not simply attempting to blindly "disrupt" or change them. Let's build inputs for today's user and outputs for today's doctor, not some future fantasy of magically disrupted clinical routines and practices.

3. Design that matters for patients, not just consumers: Markets are great at solving certain kinds of problems, and rapid innovation is happening in profit-driven consumer-facing health and wellness data. But data need to be interpreted across multiple social worlds, and designers of technologies and applications for wellness consumers will have enormous influence over what counts as actionable data in the social world of regulated health care. Designing for patients, not just consumers, means these data might more readily, and ethically, be able to bridge different worlds. Let's build technologies and data solutions that create transparent, transportable data that can be as useful to doctors in

“POLICY MAKERS, ADVOCATES, AND TECHNOLOGY DESIGNERS ALIKE MUST REMEMBER THAT THE SOLUTIONS FOR THE PROBLEMS OF HEALTH INFORMATION INNOVATION ARE AS MUCH SOCIAL AS THEY ARE TECHNICAL.”

clinical decision making as they are to individuals at improving their lives.

4. New models for patient–doctor communication: A doctor is one expert among many different kinds of expertise, information, data, and knowledge. We need new models for how health-care providers can bring their expertise to patients’ data—for interpretation and clinical decision making, as well as for hearing patients’ stories through their data and understanding them quantitatively and contextually. The image that my collaborator Anthony L. Back, an oncologist who is working to change how other cancer doctors talk with their patients, uses that of doctors providing only one part of a widening information stream that patients navigate. Doctors need to be open to evaluating new kinds of information from their patients. Social media tools designed to support communication with clinical needs and uses in mind could be one step in that direction. We need to build better tools to allow patients to share more information and data with their doctors in ways that make it possible, feasible, and practical to bring medical expertise back into the conversation.
5. Policy that embraces technological innovation (but is not besotted by it): *Disruption* only goes so far as a roadmap for change. The policy conversations that begin with the requirements for electronic health records and health information exchange standards will help frame public policy on data that will have wide-reaching impact. FDA guidelines on mHealth and mobile medical apps, expected later this year, will bring clarity and stability to the field. The enthusiasm for data, though, should not overshadow the fact that the United States still needs to provide more basic and preventative health care to more people at a cheaper cost—a problem whose solution we will almost certainly not find in big data.

At last year’s Stanford Medicine X Conference, a speaker confidently gave a simple, linear equation: “Data leads to knowledge which leads to change.” This seemed sensible enough to most in the room because it reflects the values of quantified self and data-driven health innovation. An audience member, however, changed the tone of the discussion by responding, “If knowledge translated into behavior we wouldn’t need psychologists.” At the heart of many current attempts at data-driven health is a powerfully seductive but inherently flawed model of the relationship of data to knowledge, interpretation, and action. For individual users this model relies on a clear and direct relationship of information to behavior, which has been refuted by generations of psychologists (we know what is bad for us but still do it) and health communication scholars (“hearing” such messages is far from straightforward). And yet, most mobile health and wellness applications advertise their usefulness by asking potential users to trust that their data, absent of a rich context

of actions, can lead to change. For health enterprise users, this model fails to address the practices and legacies and traditions that make or break the successful adoption technology. Without consideration of social uses and clinical practices, big data will fail to cure the woes of the U.S. health-care system.

Acknowledgment

The author’s work on consumer biosensing is supported by a gift from Intel.

Author Disclosure Statement

No conflicting financial interests exist.

References

1. Ginger.io. <http://ginger.io/> (Last accessed on March 18, 2013).
2. Schneier B. The future of privacy. Available online at www.schneier.com/blog/archives/2006/03/the_future_of_p.html (Last accessed on July 5, 2013).
3. Lu S. Can we paint a personal health picture from our daily digital traces? Available online at <http://blog.tedmed.com/?p=3586> (Last accessed on July 5, 2013).
4. Swann M. The big data self: Fundamental disruption and biological discovery. *Big Data* 2013; 1:85–99.
5. Downs S. The body-data craze, the hype cycle and why it matters. Available online at www.rwjf.org/en/blogs/pioneering-ideas/2013/07/body-data_craze_.html (Last accessed on July 5, 2013).
6. Gitelman L (Ed.). “Raw Data” Is an Oxymoron. Cambridge, MA: MIT Press, 2013.
7. Neff G, Fiore-Silfvast B. Pictures of health: Does the future of wellness need us? Theorizing the Web, NYC 2013. Extended abstract available online at <http://thesocietypages.org/cyborgology/2013/02/26/ttw13-preview-gina-neff-and-brittany-fiore-silfvast-pictures-of-health-does-the-future-of-wellness-need-us/> (Last accessed on July 5, 2013).
8. Sainani K. Statistically significant: Biostatistics is blooming. *Stanford Medicine* 29:16–19, 40–41. Available online at <http://stanmed.stanford.edu/2012summer/article2.html> (Last accessed on August 9, 2013).
9. Kolata G. Accord aims to create a trove of genetic data. *New York Times*, June 6, 2013, A6.
10. See the Beacon Nation Learning Guides. Available online at www.hibeacon.org/index.php/beacon_nation/learning_guides (Last accessed on August 9, 2013).
11. Fox, S. Opening keynote address: What is the future of self tracking? *Stanford Medicine X*, September 2012. Available online at <http://vimeo.com/60148779> (Last accessed on August 9, 2013).

12. Patterson H, Nissenbaum N. n.d. Context-dependent expectations of privacy in self-generated mobile health data. Working paper. Media, Culture and Communication Department, New York University, 52 p.
13. Fox S, Duggan M. 2013. Health online 2013. Available online at <http://pewinternet.org/Reports/2013/Health-online.aspx> (Last accessed on August 9, 2013).
14. Crawford K. 2013. The hidden biases in big data. Available online at http://blogs.hbr.org/cs/2013/04/the_hidden_biases_in_big_data.html (Last accessed on August 9, 2013).
15. Sweeney L, Abu A, Winn J. 2013. Identifying Participants in the Personal Genome Project by Name. Working paper. Harvard University. Available online at [http://data](http://data.privacylab.org/projects/pgp/1021-1.pdf)
[privacylab.org/projects/pgp/1021-1.pdf](http://data.privacylab.org/projects/pgp/1021-1.pdf) (Last accessed on July 5 2013).
16. Tene O, Polonetsky J. Privacy in the age of big data: time for big decisions. *Stan. L. Rev. Online* 2012; 64:63–69.

Address correspondence to:

Gina Neff
Department of Communication
University of Washington
Box 353740
Seattle, WA 98195

E-mail: gneff@uw.edu