

# Big Data in Science and Healthcare: A Review of Recent Literature and Perspectives

## Contribution of the IMIA Social Media Working Group

M. M. Hansen<sup>1</sup>, T. Miron-Shatz<sup>2</sup>, A. Y. S. Lau<sup>3</sup>, C. Paton<sup>4</sup>

<sup>1</sup> School of Nursing and Health Professions, University of San Francisco, San Francisco, California, USA

<sup>2</sup> Center for Medical Decision Making, Ono Academic College, Kiryat Ono, Israel

<sup>3</sup> Centre for Health Informatics, Australian Institute of Health Innovation, University of New South Wales, Australia

<sup>4</sup> George Institute for Global Health, University of Oxford, Oxford, UK

### Summary

**Objectives:** As technology continues to evolve and rise in various industries, such as healthcare, science, education, and gaming, a sophisticated concept known as Big Data is surfacing. The concept of analytics aims to understand data. We set out to portray and discuss perspectives of the evolving use of Big Data in science and healthcare and, to examine some of the opportunities and challenges.

**Methods:** A literature review was conducted to highlight the implications associated with the use of Big Data in scientific research and healthcare innovations, both on a large and small scale.

**Results:** Scientists and health-care providers may learn from one another when it comes to understanding the value of Big Data and analytics. Small data, derived by patients and consumers, also requires analytics to become actionable. Connectivism provides a framework for the use of Big Data and analytics in the areas of science and healthcare. This theory assists individuals to recognize and synthesize how human connections are driving the increase in data. Despite the volume and velocity of Big Data, it is truly about technology connecting humans and assisting them to construct knowledge in new ways.

**Concluding Thoughts:** The concept of Big Data and associated analytics are to be taken seriously when approaching the use of vast volumes of both structured and unstructured data in science and health-care. Future exploration of issues surrounding data privacy, confidentiality, and education are needed. A greater focus on data from social media, the quantified self-movement, and the application of analytics to “small data” would also be useful.

### Keywords

Science, healthcare, higher education, big data, analytics, quantified self Introduction, connectivism

Yearb Med Inform 2014;21-6

<http://dx.doi.org/10.15265/IY-2014-0004>

Published online August 15, 2014

### Introduction

Currently, multiple worldwide enterprises are asking key questions about “Big Data,” which has become a buzzword. For those who are willing to listen, Big Data is offering valuable patterns and predictions in the world today. It is not surprising this concept is recently receiving a lot of attention. According to Asigra, a Cloud Backup company since 1986, a staggering 90% of the data in the world today have been created only during the last two years [1]. And, it is predicted the worldwide number of Internet Protocol (IP) addresses will quadruple by 2015. Moreover, it is forecasted three billion people will be online creating close to eight zettabytes of data two years from now [1]. This amount of data may appear alarming while at the same time interesting when companies such as Google harness personal input data and forecast flu epidemics in collaboration with the Centers for Disease Control (CDC) [2].

Besides the legacy of electronic bulletin boards and listservs we now have large volumes of data being produced by multiple users of social media platforms [3]. While Electronic Medical Records (EMRs) contain a plethora of data, such as patient demographics, clinical and genomic data, and are known for assisting with the flow of health care, today they are seen as a way for performing large-scale and low-cost health care analysis and decision-making. EMR data sharing has its challenges, such as patient privacy and, privacy has to be a

high priority in order to comply with the EU Directive 95/46/CE and the HIPAA privacy rule [4].

In regards to the increased use of Social Media tools, an example of Big Data is the fact that “32 billion searches” were performed via Twitter during the month of August 2012 [1]. Atule Butte (@atulebutte) tweets about wearable devices that assist [aspiring] fitness buffs to track their personal data [5]. As wearable devices become more popular and accepted, even for those with poor posture [5], personal quantifiable data will add to the exploding 2.5 quintillion data bytes per day [1]. The increased use of telehealth will further test the storage capacity for patient data and the innovative use of Google Glass by physicians will also add to the social and behavioral aspects of Big Data [6]. The healthcare industry has been slow to embrace Big Data due to the cost of adding analytic functions to existing EHRs, privacy issues, poor-quality data, and a lack of willingness to share data [7]. However, today more professionals are seeing the need to listen and act upon Big Data to benefit health outcomes through online communication and sharing of data. The aim of this paper is to provide the reader a glimpse of the literature centering on the challenges and opportunities in analytics of Big Data in science and health care. We begin by discussing the science of big data and the need to balance between quantity and quality, and then move on to small data and its challenges, which are a small scale reflection of the big data challenges.

## The Science of Big Data

Over the past century, scientific advances in medicine have generally been made using a “frequentist” approach to statistical analysis: Samples of populations are studied and the results from the samples are extrapolated to estimate the effects of the intervention being studied. For most types of experiment, sampling data is sufficient to build an effective picture of the entire dataset and, statistically, we can give high levels of accuracy to predictions based on relatively small samples. Data collected in this way is often of very high quality. To ensure the sample is representative and accurate, the data is collected and ‘cleaned’ with great care. This extra care is often very expensive, however, and over the last few decades we have seen the costs of running large randomized control trials spiral upwards.

Big Data offers a potential solution to this issue. Although data produced from such sources as social networking communities, EHR systems, and wearable devices are generally of much lower quality than data carefully collected by researchers looking to answer specific questions, the sheer volume of the data may outweigh their messiness. In addition, there is also a trend to higher quality ‘big data’ collection such as the data produced in genomic analysis and structured data that can be generated from standard-compliant EHR systems. As the percentage of the population being sampled approaches 100%, messy data can have greater predictive power than highly cleaned and carefully collected data that might only be a sample of 1% of the target population for the researcher [8]. The quantity of data alters the way and approaches used to relate, utilize, and understand data.

In addition to just having more data, Big Data also generally refers to the application of machine learning for analyzing the data sets. Machine learning effectively turns the scientific method on its head. Instead of researchers creating a hypothesis and collecting data from a sample of the population, machine-learning algorithms plow through large data sets searching for hypotheses. They do this through a process

of brute force classification (finding and matching clusters of correlations in the data) combined with a process of learning and feedback to make the process more efficient. Machine learning algorithms are generally quite simple and are really just looking for associations between different elements of the data.

Because of this, we need to take the results of Big Data machine learning algorithms for what they are: new hypotheses rather than firm predictions. Researchers can test the hypotheses to a limited extent by dividing the datasets or re-running the algorithms on new data collected. But to gather the best evidence on a particular question, it may still be necessary to run a prospective “frequentist”-style trial to test any strong hypothesis that come out of the machine learning process, particularly when trying to answer questions about human health.

## Healthcare Sector

While researchers are still debating the definitions and boundaries of Big Data in health, benefits of health-related Big Data have been demonstrated in three areas so far, namely to 1) prevent disease, 2) identify modifiable risk factors for disease, and 3) design interventions for health behavior change [9]. Organizations worldwide are recognizing the Big Data movement and introducing new initiatives for knowledge discovery and data-driven decision-making. For example, the National Institute of Health (NIH) is establishing the Big Data to Knowledge (BD2K) and Infrastructure Plus Program, which provides a shared computational environment (e.g. data standards, ontologies, data catalogues, virtualized cloud computing) to facilitate large-scale biomedical data analysis for the NIH community [10]. Specifically, the NIH US Library of Medicine hosts an impressive set of data sharing repositories [11], which primarily accept submissions of biomedical data and other information sharing systems from NIH-funded investigators. In addition, the United Nations (UN) is

launching the Global Pulse project, which advocates for the ‘data philanthropy’ movement by asking organizations and individuals to contribute data, resources, and skills to help understand the impact of UN development programs and ways to improve their outreach on affected populations and regions [12].

Big Data streams in health can be broadly summarized into three categories [13]. Traditional medical data is primarily originated from the health system (e.g. EMRs, personal and family health history, medication history, lab reports, pathology results), where the objective of these analyses is to derive a better understanding of disease outcomes and their risk factors, reduce health system costs, and improve its efficiency [13]. “Omics” data refer to large-scale datasets in the biological and molecular fields (e.g. genomics, microbiomics, proteomics, and metabolomics), where the aim of these analyses is to understand the mechanisms of diseases and accelerate the individualization of medical treatments (e.g. “precision medicine”) [3,6]. As pointed out by Alice Whittmore, in the Stanford Big Data in Biomedicine Conference (2013), genomic testing and mapping could, for example, point to women in high risk of developing breast cancer, which would allow allocating them preventive care, and reduce the need for large scale, potentially hazardous interventions, for other low-risk women [14]. Last but not least, data from social media and the quantified-self movement essentially consist, of signs and behaviors on how individuals (or groups of individuals) use the Internet, social media, mobile applications (apps), sensor devices, wearable computing devices, or other technological and non-technological tools to better inform and enhance their health.

This section presents examples of health-related Big Data projects, with an emphasis on data from social media and the quantified-self movement (Table 1). For big data research related to EMRs, digital enterprise, genetic data and omics sources, readers can refer to the following reviews and perspectives conducted recently [15, 16, 17, 18, 19].

**Table 1:** Examples of health-related Big Data projects related to social media and the quantified-self movement [7, 56, 58, 13].

Data type	How has it been used in health?	Examples
Quantified-self data (via devices, self-reporting, or sensors)	<ul style="list-style-type: none"> <li>Engaged in the self-tracking of signs and/or behaviors as n=1 individual or in groups, where there is often a proactive stance toward acting on the information [13]</li> <li>Provides richer and more detailed data on potential risk factors (biological, physical, behavioral or environmental) [13]</li> <li>Allows data collection over potentially longer follow-up periods than is currently possible using standard questionnaires [13]</li> </ul>	<ul style="list-style-type: none"> <li>Food consumption [20]</li> <li>Information diet [21]</li> <li>Smile triggered electromyogram (EMG) muscle to create unexpected moments of joy in human interaction [22]</li> <li>Coffee consumption, social interaction, and mood [23]</li> <li>Idea-tracking process [24]</li> <li>Use of rescue and controller asthma medications with an inhaler sensor (e.g. Asthmapolis) [25]</li> <li>Monitors blood glucose levels in diabetics (e.g. Glooko) [26]</li> <li>Psychological, mental and cognitive states and traits (e.g. MyCompass) [27]</li> <li>Physical activity (e.g. FitBit; Jawbone Up, RunKeeper) [28, 29, 30]</li> <li>Diet (e.g. My Meal Mate) [31]</li> <li>Sleep quality (e.g. Lark) [32]</li> <li>Medication adherence (e.g. MyMedSchedule) [33]</li> </ul>
Location-based information	<ul style="list-style-type: none"> <li>Information derived from Global Positioning Systems (GPS), Geographic Information Systems (GIS), and other open source mapping and visualization projects</li> <li>Provides information on the environmental and social determinants of health</li> <li>Monitors for disease outbreaks near your location</li> </ul>	<ul style="list-style-type: none"> <li>Weather patterns, pollution levels, allergens, traffic patterns, water quality, walkability of neighborhood, and access to fresh fruit and vegetables (such as supermarkets) [34, 35, 36]</li> <li>HealthMap [37]</li> </ul>
Twitter (Note: a 2011 study has suggested that 8.5% of English-language tweets relate to illness, and 16.6% relate to health [46])	<ul style="list-style-type: none"> <li>Assesses disease spread in real-time</li> <li>Assesses sentiments and moods</li> <li>Facilitates emergency services by allowing for the wide-scale broadcast of available resource, enabling people in need of medical assistance to locate help</li> <li>Facilitates crisis mapping (e.g. where eyewitness reports are plotted on interactive maps. These data can help target areas for emergency services and additional resources)</li> <li>Facilitates discourse on non-emergency health-care (e.g. broadcasts of public health messages, quantify medical misconception)</li> </ul>	<ul style="list-style-type: none"> <li>Quantify medical misconceptions (e.g. concussions) [38]</li> <li>The spread of poor medical compliance (e.g., antibiotic use) [39]</li> <li>Trends of cardiac arrest and resuscitation communication [40]</li> <li>Cervical and breast cancer screening [41]</li> <li>Postpartum depression [42]</li> <li>Influenza A H1N1 outbreak (disease activity and public concern) [43]</li> <li>2010 Haitian cholera outbreak [44]</li> <li>Emergency situations from Boston marathon explosion [45]</li> </ul>
Health-related social networking sites	<ul style="list-style-type: none"> <li>Facilitates sharing of personal health data and advice amongst patients and consumers</li> <li>Monitors spread of infectious diseases via crowd surveillance</li> </ul>	<ul style="list-style-type: none"> <li>PatientsLikeMe [47]</li> <li>Disease surveillance sites which collect participant-reported symptoms and utilize informal online data sources to analyze, map, and disseminate information about infectious disease outbreaks (e.g. Flu Near You, HealthMap, GermTracker, Sickweather) [37, 48, 49, 50]</li> </ul>
Other social networking sites (e.g. online discussion board, Facebook)	<ul style="list-style-type: none"> <li>Monitors how patients use social media to discuss their concerns and issues</li> <li>Provides awareness of what the “person in the street” is saying [56]</li> </ul>	<ul style="list-style-type: none"> <li>Side effects and associated medication adherence behaviors (e.g. drug switching and discontinuation) [51]</li> </ul>
Search queries and Web logs	<ul style="list-style-type: none"> <li>Found to be highly predictive for a wide range of population-level health behaviors</li> <li>Search keyword selection has been found to be critical for arriving at reliable curated health content</li> <li>“Click” stream navigational data from web logs are found to be informative of individual characteristics such as mental health and dietary preferences [57]</li> </ul>	<ul style="list-style-type: none"> <li>Google and Yahoo search queries have been used to predict epidemics of illnesses, such as: <ul style="list-style-type: none"> <li>Influenza (Google 2013)</li> <li>Dengue fever [52]</li> <li>Seasonality of mental health, depression and suicide [53]</li> <li>Prevalence of Lyme disease [54]</li> <li>Prevalence of smoking and electronic cigarette use [55]</li> </ul> </li> </ul>

## Small Data – Do patients Make Sense of Their Data and Use It to Improve Health?

While the paper focuses on Big Data, this section focuses on how patients (and people in general) use the small, personal data that is generated on their personal apps and tracking devices. Indeed, “the quantified self is a natural progression from the current practice of the patient being monitored by health professionals to individuals monitoring themselves” [59]. Some have identified a trend of “citizen science,” in which non-professionally trained individuals conduct science-related activities [60]. This begs the question of whether self-monitoring, and informed use of tracking information by patients - not to mention the ability to become a mini-expert, identifying trends, and acquiring specialized, quasi-scientific knowledge of one’s disease or condition - are prevalent and easily obtainable. Several issues, known from psychological research, suggest obtaining this goal is far from trivial.

Primarily, to use data, one first has to make sense of it. Yet comprehension cannot be taken for granted. Studies examining how people understand simple probabilistic information pertaining to prostate or breast cancer have found mistake rates to hover around 50% [61, 62]. Furthermore, miscomprehension also occurred when students were presented with information on prenatal testing [63]. This suggests whatever data or trends we expect patients to benefit from, need to be tested for clarity and understandability with low health literacy taken into account [64]. Comprehension is further hindered when people, physicians included, are presented with more than 3 pieces of information at a time [65]. In addition, once one has made sense of data, one also needs to be motivated to change the behavior. An interesting case comes from the Federal Drug Administration (FDA) warning the administration of cough and cold medication to children under the age of two. A comparison of experienced parents, who had raised children over the critical age of two, and inexperienced parents found more than half (53.3%) of inexperienced parents adhered to the FDA warning,

compared with just over a quarter (28.4%) of experienced parents [66]. The researchers concluded that experience, such as having given a child cough and cold medication numerous times in the past, with no adverse effects, was more influential than information delivered through a warning. In the context of using one's own data to improve one's health, it might be tracking health indicators, especially if routinely performed, will serve as actual experience and will motivate human action.

However, comprehension of information, and motivation for change, are not always enough. Patients required to detect a change, for example, those following repeated measurements of their blood sugar levels, may not always know what to do in order to reduce it: Should they change their medication? Eat differently? Exercise more? This is where a healthcare professional's involvement is called for. And Big Data provides just the opportunity. As Kim [67] suggests: people involved in the quantified self movement will still want to share information with their physicians and healthcare providers. That way they can receive better, more personalized care that is based on their health conditions, diet, and level of physical activity [67]. Just like Feinberg [68] reminds us, patients may wish to have varying degrees of involvement in the treatment process, and, we can extrapolate – patients may have varying degrees of ability to determine the required course of action based on self-tracked information. Yet, apps, devices, and wearables are for the most part sold to consumers, regardless of the physician's awareness or input. Not only are the physicians unaware the tracking device was purchased, interoperability, legal, and privacy issues may prevent healthcare professionals from approaching this data or making use of it.

A recent attempt to help patients integrate input from various self-tracking sources, to make sense of it, and even to connect it to medical records, comes from a US insurer, Aetna, which developed an access-free platform for such integration. While everyone can use the platform, only Aetna members have access to their medical information [69]. Reservations aside, track-

ing devices, apps, and other means of collecting patient and consumer input, have the potential to empower and inform patients, as well as to advance science. In some cases, this happens through patient participation in online and other data collection endeavors, such as the ones on PatientsLikeMe [47], a website inviting patients to monitor their disease and share data so knowledge is accumulated regarding their condition. For example, amyotrophic lateral sclerosis (ALS) patients reported their use of limbs and associated it with disease onset, which allowed for the identification of trends in onset. This detailed information may not have been available otherwise [70]. Patient partnership in entering the data in a personal health record (PHR), and the ownership they feel of the process that may happen in their home and is controlled by them, rather than by a health professional, may assist in introducing greater trust.

If a recommendation is generated based on a patient's personal data, it might be perceived as better suited to them, trustworthier, and the patient will be more likely to act accordingly. This may help circumvent the issue of relatively low trust in government health agencies such as the FDA [71] as opposed to far greater trust in, for example, one's pediatrician, who of course you know in person [66]. Patients may derive additional benefits from reporting and tracking their medical data, benefits that may be different from developing an expertise in their disease. For example, patients benefit from the ability to know how well they are doing in comparison with others [72]. And patients who reported their symptoms and other personal health information on PatientsLikeMe reported an increased comfort in sharing such information [73]. Notably, this does not require comprehending the meaning and trends in one's information. It comes from the mere opportunity to share one's data, and to have it accepted without judgment. It may translate to these patients feeling more secure and being more open when discussing their condition outside the realm of the health social network. This suggests small data is beneficial to patients on many levels, which may be quite different than the Big Data angle.

## Connectivism

The connectivist approach takes ideas from brain models and neural networks in learning from technologies [74]. Therefore, a few of the principles related to connectivism are that learning may reside in machines, maintaining connections is necessary to create constant learning and, up-to-date knowledge is the core of connectivist learning moments. Connectivism as an analogy to health is evident. Health requires not only knowledge but also a connected relationship between the provider and the patient, and personalization such that interventions are tailored to the patient's unique preferences and form of conduct, such as drug adherence. Different people have different reasons for non-adherence to medications [75]. Furthermore, connectivism may serve as an underlying theory for how massive amounts of data collected through various technologies, connect humans and afford interactions in science, healthcare, and education. Hussain [76] explored the underlying principles associated with Siemens' [74] connectivist theory of learning that is historically considered the go-to theory supporting learning in the digital age. Hussain posits connectivism may need to be reconsidered in the advent of "ambient mobile pervasive communication (p.14)" consisting of filtering mechanisms and smart agents. And, this query has been investigated with an overarching suggestion connectivism still remains a strong theory for understanding Big Data and its initial links to human interactions with technology.

## Concluding Thoughts

Recognizing, understanding, and using Big Data in terms of scientific research and healthcare are necessary at this time in order to arrive at best evidence in a world of ever increasing data. Further investigation into the limitations of Big Data, such as inconsistencies regarding standards, policy, ethics, gaps in structured databases and finding a way to contain and deliver Big Data in a meaningful way to health care practitioners is interesting and necessary. This review presents just a glimpse of current and cogent literature



illustrating and supporting the use of Big Data in two areas. Another area to consider is education because of online education and today's classroom milieu-ubiquitous powerful mobile learning devices becoming more mainstream. The fascinating concepts of Big Data and analytics are not to be ignored in this unprecedented era of innovative technologies that create colossal volumes of both structured and unstructured data. Future papers directed at issues surrounding the open problem of "Quo vadis" (data privacy), confidentiality, and learning analytics are needed. The confluence of Big Data interpretations will continue given the proliferation of data from scientific led endeavors, accelerating healthcare innovations, and the rise of Big Data in higher education as a result of embedding technologies and the proliferation of e-Learning in higher education.

### Acknowledgments

The lead author would like to sincerely thank co-authors Talya Miron-Shatz, Annie Lau, and Chris Paton for their timely contributions to this chapter in light of their very busy schedules. Annie Lau was supported by National Health and Medical Research Council (NHMRC) Centre of Research Excellence in Informatics and E-Health (1032664).

### Conflict of Interest

The authors declare no conflicts of interest.

### References

- Rasid Z. Big data infographic: What is big data? [Internet] 2013. [Cited March 17, 2014]. Available from: <http://www.asigra.com/blog/big-data-infographic-what-big-data>.
- Dugas AF, Hsieh YH, Levin SR, Pines JM, Mariniss DP, Moharebet al. Google Flu Trends: correlation with emergency department influenza rates and crowding metrics. *Clin Infect Dis* 2012;54(4):463-9.
- Smith C. How many people use the top social media? [Internet] 2011. [Cited March 17, 2014]. Available from: <http://expandedramblings.com/index.php/resource-how-many-people-use-the-top-social-media/#.UxvcnuddWL>.
- Gkoulalis-Divanis A, Loukides G. Privacy Challenges and Solutions for Medical Data Sharing. Zurich. Medical privacy tutorial. [Internet] 2011. [Cited March 17, 2014]. Available from: <http://www.zurich.ibm.com/medical-privacy-tutorial/>
- Butte A. Wearable devices getting more niche. [Internet] 2014. [http://blogs.wsj.com/digits/2014/03/04/wearable-devices-getting-more-niche/?utm\\_content=buffer7044b&utm\\_medium=social&utm\\_source=twitter.com&utm\\_campaign=buffer&mg=blogs-wsj&url=http%253A%252F%252Fblogs.wsj.com%252Fdigits%252F2014%252F03%252F04%252Fwearable-devices-getting-more-niche%253Futm\\_content%253Dbuffer7044b%2526utm\\_medium%253Dsocial%2526utm\\_source%253Dtwitter.com%2526utm\\_campaign%253Dbuffer](http://blogs.wsj.com/digits/2014/03/04/wearable-devices-getting-more-niche/?utm_content=buffer7044b&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer&mg=blogs-wsj&url=http%253A%252F%252Fblogs.wsj.com%252Fdigits%252F2014%252F03%252F04%252Fwearable-devices-getting-more-niche%253Futm_content%253Dbuffer7044b%2526utm_medium%253Dsocial%2526utm_source%253Dtwitter.com%2526utm_campaign%253Dbuffer)
- Paton C, Will Google Glass become standard eyewear for Doctors? [Internet] 2014. [Cited March 17, 2014]. Available from: [http://www.healthinformaticsforum.com/forum/topic/show?id=2068976%3ATopic%3A149324&xgs=1&xg\\_source=msg\\_share\\_topic](http://www.healthinformaticsforum.com/forum/topic/show?id=2068976%3ATopic%3A149324&xgs=1&xg_source=msg_share_topic)
- O'Connor F. Health-IT early adopters well poised for big-data advances in clinical medicine. [Internet] 2013. [Cited March 17, 2014]. Available from: [http://www.computerworld.com/s/article/9238063/Health\\_IT\\_early\\_adopters\\_well\\_poi](http://www.computerworld.com/s/article/9238063/Health_IT_early_adopters_well_poi)
- Mayer-Schönberger V, Cukie K. Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt; 2013. Chapter 2.
- Barrett MA, Humblet O, Hiatt RA, Adler NE. Big data and disease prevention: From quantified self to quantified communities. *Big Data* 2013;1(3):168-75.
- Coakley MF, Leerkes MR, Barnett J, Gabriellian AE, Noble K, Weber MN, et al. Unlocking the power of big data at the national institute of health. *Big Data* 2013;1(3).
- NIH Data Sharing Repositories [Internet] 2014. [Cited March 5, 2014]. Available from: [http://www.nlm.nih.gov/NIHbmic/nih\\_data\\_sharing\\_repositories.html](http://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html).
- Kirkpatrick R. Big data for development. *Big Data* 2013;1(1):3-4.
- Swan M. The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data* 2013 Jun;1(2):85-99.
- Whittemore A. Improving Health Systems with Big Data. Big Data in Biomedicine conference. Stanford School of Medicine. [Internet] 2013. [Cited March 18, 2014]. Available from: [https://mediaspace.stanford.edu/media/Improving+Health+Systems+with+Big+Data/0\\_vyo4glde](https://mediaspace.stanford.edu/media/Improving+Health+Systems+with+Big+Data/0_vyo4glde)
- Bourne PE. What Big Data means to me. *J Am Med Inform Assoc* 2014;21(2):194.
- Kum H-C, Krishnamurthy A, Machanavajhala A, Reiter MK, Ahalt S. Privacy preserving interactive record linkage (PPIRL). *J Am Med Inform Assoc* 2014;21(2):212-20.
- Shivade C, Raghavan P, Fosler-Lussier E, Embe PJ, Elhadad N, Johnson SB, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. *J Am Med Inform Assoc* 2014 Mar-Apr;21(2):221-30.
- Shoenbill K, Fost N, Tachinardi U, Mendonca EA. Genetic data and electronic health records: a discussion of ethical, logistical and technological considerations. *J Am Med Inform Assoc* 2014;21(1):171-80.
- Tenenbaum JD, Sansone S-A, Haendel M. A sea of standards for omics data: sink or swim? *Journal of the American Medical Informatics Association*. 2014;21(2):200-3.
- Yau N. A year of food consumption visualized. FlowingData [Internet] 2011 [Cited March 17, 2014]. Available from: <http://flowingdata.com/2011/06/29/a-year-of-food-consumption-visualized/>
- McCormick T. Video of my healthier information talk. [Internet] 2012. [Cited November 29, 2013]. Available from: <http://tjm.org/2012/04/17/videoof-my-healthier-information-talk/>
- Tasse D. Quantified self 2012: Some cool things. Tales 'n' ideas. [Internet] 2012. [Cited November 29, 2013]. Available from: [http://talesnideas.blogspot.com/2012\\_09\\_01\\_archive.html](http://talesnideas.blogspot.com/2012_09_01_archive.html)
- Wolf G. The data-driven life. *The New York Times*. [Internet] 2012. [Cited November 29, 2013]. Available from: <http://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html>
- Chua S. Notes from the quantified self 2012 conference. [Internet] 2012. [Cited November 29, 2013]. Available from: <http://sachachua.com/blog/p/23723/>
- Asthmapolis. [Internet] 2013. [Cited November 29, 2013]. Available from: <http://propellerhealth.com/>
- Glooko. [Internet] 2013. [Cited November 29, 2013]. Available from: <http://www.glooko.com/>
- Harrison V, Proudfoot J, Wee PP, Parker G, Pavlovic DH, Manicavasagar V. Mobile mental health: review of the emerging field and proof of concept study. *J Ment Health* 2011;20(6): 509-24.
- FitBit. [Internet] 2013. [Cited November 29, 2013]. Available from: <http://www.fitbit.com/>
- JawboneUp. [Internet] 2013. [Cited November 29, 2013]. Available from: <https://jawbone.com/up>
- RunKeeper. [Internet] 2013. [Cited November 29, 2013]. Available from: <http://runkeeper.com/>
- Carter MC, VJ Burley, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: Pilot randomized controlled trial. *J Med Internet Res* 2013;15(4): e32.
- Lark. [Internet] 2013. [Cited November 29, 2013]. Available from: <http://lark.com/>
- Dayer L, Heldenbrand S, Anderson P, Gubbins PO, Martin BC. Smartphone medication adherence apps: Potential benefits to patients and providers. *J Am Pharm Assoc* 2003;53(2):172-81.
- de Nazelle A, Seto E, Donaire-Gonzalez D, Mendez M, Matamala J, Nieuwenhuijsen MJ, et al. Improving estimates of air pollution exposure through ubiquitous sensing technologies. [Internet] 2013. [Cited November 30, 2013]. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23416743>
- Cerin E, Lee KY, Barnett A, Sit CH, Cheung MC, Chan WM. Objectively-measured neighborhood environments and leisure-time physical activity in Chinese urban elders. *Prev Med* 2013;56(1):86-9.
- Robinson PL, Dominguez F, Teklehaimanot S, Lee M, Brown A, Goodchild M. Does distance decay modeling of supermarket accessibility predict fruit and vegetable intake by individuals in a large metropolitan area? *J Healthcare Poor Underserved* 2013;24(1 Suppl):172-85.
- HealthMap. [Internet] 2013 [Cited March 18, 2014]. Available from: <http://healthmap.org/>
- Sullivan SJ, Schneiders AG, Cheang CW, Kitto

- E, Lee H, Redhead J, et al. What's happening? A content analysis of concussion-related traffic on twitter. *Br J Sports Med* [Internet] 2011. [Cited November 30, 2013]. Available from: [http://www.academia.edu/1037228/Whats\\_happening\\_A\\_content\\_analysis\\_of\\_concussion-related\\_traffic\\_on\\_Twitter](http://www.academia.edu/1037228/Whats_happening_A_content_analysis_of_concussion-related_traffic_on_Twitter)
39. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: twitter and antibiotics. *Am J Infect Control* 2010;38(3): 182-8.
40. Bosley JC, Zhao NW, Hill S, Shofer FS, Asch DA, Becker LB, et al. Decoding twitter: Surveillance and trends for cardiac arrest and resuscitation communication. *Resuscitation*. 2013;84(2):206-12.
41. Lyles CR, Lopez A, Pasick R, Sarkar U. 5 mins of uncomfyness is better than dealing with cancer 4 a lifetime: an exploratory qualitative analysis of cervical and breast cancer screening dialogue on twitter. *J Cancer Educ* 2013;28(1):127-33.
42. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. 2013 ACM Annual Conference on Human Factors in Computing Systems. 2013: 3267-76.
43. Chew C, Eysenbach G. Pandemics in the age of twitter: Content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One* 2010;5(11): e14118.
44. Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *Am J Trop Med Hyg* 2012;86(1): 39-45.
45. Cassa CA, Chunara R, Mandl K, Brownstein JS. Twitter as a sentinel in emergency situations: lessons from the Boston marathon explosions. *PLoS Curr* 2013;5
46. Paul MJ, Dredze M. A Model for Mining Public Health Topics from Twitter. Baltimore: Technical report. Johns Hopkins University, 2011.
47. PatientsLikeMe. [Internet] 2013. [Cited November 30, 2013]. Available from: <http://www.patientslikeme.com/>
48. FluNearYou. [Internet] 2013. [Cited November 30, 2013]. Available from: <https://flunearyou.org/>
49. GermTracker. [Internet] 2013. [Cited November 30, 2013]. Available from: <http://germtracker.org/>
50. Sickweather. [Internet] 2013. [Cited November 30, 2013]. Available from: <http://www.sickweather.com/>
51. Mao JJ, Chung A, Benton A, Hill S, Ungar L, Leonard CE, et al. Online discussion of drug side effects and discontinuation among breast cancer survivors. *Pharmacoepidemiol Drug Saf* 2013;22(3): 256-62.
52. Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *PLoS Negl Trop Dis* 2011;5(5): e1206.
53. Ayers J W, Althouse BM, Allem JP, Rosenquist JN, Ford DE. Seasonality in seeking mental health information on Google. *Am J Prev Med* 2013;44(5): 520-5.
54. Seifter A, Schwarzwald A, Geis K, Aucott J. The utility of „Google Trends“ for epidemiological research: Lyme disease as an example. *Geospat Health* 2010; 4(2):135-7.
55. Ayers JW, Ribisl KM, Brownstein JS. Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. *Am J Prev Med* 2011;40(4): 448-53.
56. Hill S, Merchant R, Ungar L. Lessons Learned About Public Health from Online Crowd Surveillance. *Big Data* September 2013;1(3):160-7.
57. West R, White RW, Horvitz E. From cookies to cooks: Insights on dietary patterns via analysis of web usage logs. 22nd International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2013. p. 1399-410.
58. Neff G. Why BD Won't Cure Us. *Big Data* September 2013;1(3): 117-23.
59. Luciano SJ, Cumming GP, Wilkinson MD, Kahana E. The emergent discipline of health web science. *J Med Internet Res* 2013 August;15(8): e166.
60. Swan M. Crowdsourced health research studies: An important emerging complement to clinical trials in the public health research ecosystem. *J Med Internet Res* 2012 Mar-Apr;14(2): e46.
61. Hanoch Y, Rolison J, Miron-Shatz T. What do men understand about lifetime risk following genetic testing? The effect of context and numeracy. *Health Psychology* 2011;31(4): 530-3.
62. Hanoch Y, Miron-Shatz T, Himmelstein M. Genetic testing and risk interpretation: How do women understand lifetime risk results? *Judgment and Decision Making* 2012;5(2): 116-23.
63. Miron-Shatz T, Hanoch Y, Graef D, Sagi M. Presentation format, numeracy, and emotional reactions: The case of prenatal screening tests. *J Health Commun* 2009;14(5):439-50.
64. Miron-Shatz T, Bowen B, Diefenbach M, Goldacre B, Mühlhauser I, Smith RSW, et al. From blind acceptance to active inquiry: Jumping the barriers to health literacy. In: Gigerenzer G, Gray JAM, editors. *Better doctors, better patients, better decisions: Envisioning healthcare 2020*. Strüngmann Forum Report (Vol. 6). Cambridge: MIT Press; 2011. p. 191-212.
65. Hanoch Y, Miron-Shatz T, Cole H, Himmelstein M, Federman AD. Choice, numeracy and physicians-in-training performance: The case of Medicare Part D. *Health Psychol* 2010;29(4): 454-9.
66. Miron-Shatz T, Barron G, Hanoch Y, Gummerman M, Doniger GM. To give or not to give: A word to the wise, two to the wiser: Parental experience and adherence to the food and drug administration warning about over-the-counter cough and cold medicine usage. *Judgment and Decision Making* 2010;5(6):428-36.
67. Kim J. The intersection of the quantified self movement and big data. [Internet] 2013. [Cited March 18, 2014]. Available from: <http://searchhealthit.techtarget.com/opinion/The-intersection-of-the-quantified-self-movement-and-big-data>
68. Feinberg HV. From shared decision making to patient-centered decision making. *Isr J Health Policy Res* 2012;1:6.
69. Aetna: Carepass. [Internet] 2013. [Cited November 30, 2013]. Available from: <https://www.carepass.com/carepass/getstarted>
70. Turner MR, Wicks P, Brownstein CA, Massagli MP, Toronjo M, Talbot K, et al. Concordance between site of onset and limb dominance in amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry* 2011 Aug;82(8):853-4.
71. Hanoch Y, Gummerum M, Miron-Shatz T. Trust and adherence to the FDA warning regarding cough and cold medicine for children under two. *Child Care Health Dev* 2010;36(6):795-804.
72. Brownstein CA, Wicks P. The potential research impact of patient reported outcomes on osteogenesis imperfecta. *Clin Orthop Relat Res* 2010 Oct;468(10):2581-5.
73. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. *J Med Internet Res* 2010;12(2):e19. [Cited November 30, 2013]. Available from: <http://www.jmir.org/2010/2/e19/v12i2e19>
74. Siemens G. [Internet] 2004. [Cited November 30, 2013]. Available from: [Connectivism: A learning theory for the digital age. http://www.elearnspace.org/Articles/connectivism.htm](http://www.elearnspace.org/Articles/connectivism.htm)
75. McHorney CA. The Adherence Estimator: a brief, proximal screener for patient propensity to adhere to prescription medications for chronic disease. *Curr Med Res Opin* 2009;25(1):215-38.
76. Hussain F. E-learning 3.0 = e-learning 2.0 + web 3.0? IADIS International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2012).

Correspondence to  
Margaret Hansen  
School of Nursing and Health Professions  
University of San Francisco  
San Francisco, California  
USA  
E-mail: [maggie2hansen@gmail.com](mailto:maggie2hansen@gmail.com)