REVIEW

BIG DATA AND DISEASE PREVENTION:

From Quantified Self to Quantified Communities

Meredith A. Barrett,[1,2] Olivier Humblet,[1,2]* Robert A. Hiatt,[3] and Nancy E. Adler[1]*

Abstract

Big data is often discussed in the context of improving medical care, but it also has a less appreciated but equally important role to play in preventing disease. Big data can facilitate action on the modifiable risk factors that contribute to a large fraction of the chronic disease burden, such as physical activity, diet, tobacco use, and exposure to pollution. It can do so by facilitating the discovery of risk factors for disease at population, subpopulation, and individual levels, and by improving the effectiveness of interventions to help people achieve healthier behaviors in healthier environments. In this article, we describe new sources of big data in population health, explore their applications, and present two case studies illustrating how big data can be leveraged for prevention. We also discuss the many implementation obstacles that must be overcome before this vision can become a reality.

Introduction

The United States faces major health challenges. Despite spending more on healthcare per person than does any other nation, the United States scores poorly on key health indicators.[1] Up to half of all deaths can be attributed to behavioral factors such as tobacco, diet, physical activity, alcohol and drug use, as well as the physical and social environment.[2,3] Preventable chronic diseases are now the most common causes of premature death. Currently, 10% of Americans rate their health as fair or poor, and 36% of adults are considered obese.[4] Over the next 20 years, Americans are projected to suffer from as many as 8.5 million new cases of diabetes, 7.3 million cases of heart disease and stroke, and over 660,000 cases of cancer, potentially costing up to $66 billion per year.[5] New approaches to research and interventions aimed at the preventable causes of these diseases will be needed to reduce the disease burden and the resulting cost.[6]

Discussions of how the accumulation of new digital information—commonly referred to as big data—will affect health have primarily revolved around the potential impact on healthcare,[7,8] and on discoveries at the molecular level for the treatment of disease.[9] Less attention has been paid to how big data could contribute to more effective disease prevention, specifically by facilitating action on the preventable risk factors that contribute to a large fraction of the chronic disease burden. We see this action as necessary on both individual and community scales. Momentum has been growing around the concept of the quantified self, in which individuals deploy sensors and monitoring devices to measure and improve their own health and behavior. This concept can be expanded and aggregated to a population level, leading to quantified communities that measure the health and activities of their population and institutions, thereby improving collective health with a data-driven approach.[10]

[1]Center for Health and Community and [3]Department of Epidemiology and Biostatistics, University of California–San Francisco, San Francisco, California.
[2]School of Public Health, Berkeley, University of California–Berkeley, Berkeley, California.

*These authors contributed equally to this work.

DOI: 10.1089/big.2013.0027

In this article we explore two ways in which big data can facilitate efforts to prevent disease: first, by increasing the capacity to understand the behavioral, social, and environmental determinants of health in populations; and second, by enabling disease prevention efforts to be better targeted towards the subpopulations for which they will be most effective, an effort which we call precision prevention. This approach builds upon the 2011 National Research Council report advocating for precision medicine,[9] which would create and analyze massive databases linking electronic health records (EHRs) to molecular data in order to accelerate research on mechanisms of disease and individualization of medical treatments. We believe this framework would be enriched by expanding the focus to include disease prevention, and incorporating data about key behavioral, social, and environmental risk factors for disease in populations.

To that end, we describe new sources of big data in population health, including new technologies that allow data collection on a much larger and faster scale than was previously possible. Two case studies describe how big data could be leveraged to impact health by increasing physical activity as well as improving asthma outcomes. We close by reviewing obstacles that remain before big data can be harnessed to achieve precision prevention. Here we examine how big data can satisfy the need for more diverse, richly contextual, and real-time data collection in population health.

“HERE WE EXAMINE HOW BIG DATA CAN SATISFY THE NEED FOR MORE DIVERSE, RICHLY CONTEXTUAL, AND REAL-TIME DATA COLLECTION IN POPULATION HEALTH.”

What Is Big Data?

Big data—the term describing the accumulation of new data from sources such as online personal activity, commercial transactions, and sensor networks—is characterized by high-volume, high-variety, and high-velocity information.[11,12] The collection, analysis and application of big data related to health is a component of a growing field that has been referred to by multiple terms, including e-health, m-health, digital health, health information technology, health 2.0, e-medicine, and many other terms in the last few years.[13]

Leveraging Big Data to Prevent Disease

Disease prevention requires two steps. *Research* first identifies modifiable risk factors for disease (e.g., diet, exercise, smoking, alcohol consumption, and environmental pollution). These insights then lead to *interventions* to ameliorate these risk factors and improve health. Public health is the discipline most engaged in deliberate disease prevention. However, traditional public health data is not high volume, high variety, or high velocity. In fact, many aspects of public health could be considered data-poor, due to modest study sample sizes, a lack of geographically linked data, and temporal lags due to lengthy data collection and dissemination cycles. Among large, longitudinal studies, participant attrition is often high and data collection can be expensive, making long-term follow-up difficult.[14,15] Big data can play a key role in both research and intervention activities and accelerate progress in disease prevention and population health.

The technological underpinning of health-focused big data is the use of sensors and smartphones to track various aspects of health and health behaviors. People are increasingly interested in tracking their health through mobile health sensors and applications, and have the requisite technology experience to do so.[16] According to a Pew report on the social life of health information, 27% of internet users age 18 and older track their own health data online, with 15% having tracked their weight, diet, or exercise routine and 17% having tracked any other health indicators online.[17] Another Pew report found that 29% of American adults who download apps to their smartphones have downloaded an app that helps them track or manage their health.[18]

The number of mobile health application users is growing rapidly and is expected to reach 247 million users by the end of 2012.[19] Disparities in smartphone and technology access are important to address, however, when considering their utility in health data collection and intervention. The most recent Pew report found that 91% of all adults in the United States own a cell phone, with 56% owning a smartphone.[20] When segmented by race, 53% of whites, 60% of Hispanics, and 64% of African American adults own smartphones. While only 43% of those earning less than $30,000 a year own a smartphone, ownership increases to 77% when considering those under age 30 in this income group.

Responding to this interest, the market for personal sensors, health applications and their combination has rapidly expanded. As smartphones have become widespread—and, for some, indispensable—in modern life,[21] they can serve as important passive and manual data collection devices. Projections estimate that 50 billion devices will connect to the internet in the next 10 years, generating 40-fold the current amount of global personal data.[22] Passive data collection through the phone’s own accelerometer and other sensors make data collection automatic and effortless. The velocity, variety, and volume of these new big data sources make them particularly relevant to both health research and interventions.

Research to identify modifiable risk factors for disease

Identifying modifiable risk factors for disease requires datasets that include information about health outcomes as well as about potential risk factors. Correlations between risk factors and disease can be identified by applying analytics on these datasets. One could ask why we need big data since this process already occurs in standard epidemiologic studies. We see two primary ways in which big data can uniquely accelerate and enrich the discovery of new risk factors for disease.

First, massive datasets allow not only population-level analyses, but also subpopulation- and personal-level analyses. Such datasets enable the discovery of personalized risk factors, which take into account the various additional variables that might confer susceptibility or resistance to a given risk factor. Identifying personalized risk factors holds the promise of giving people more effective information about how to prevent disease, and doing so in a way that is more compelling for them to act upon because it is targeted to them specifically as opposed to the "average person." For example, a certain dietary nutrient may be beneficial to some people but harmful to others, yielding an average effect that is null. Only with a sufficiently large dataset is there adequate power to detect such statistical interactions, which will yield nutritional advice that varies by person.

"THE NUMBER OF MOBILE HEALTH APPLICATION USERS IS GROWING RAPIDLY AND IS EXPECTED TO REACH 247 MILLION USERS BY THE END OF 2012."

Second, new passive sensors (e.g., for physical activity or sleep) can allow collection of richer, more detailed data on potential risk factors over longer periods of follow-up than is currently possible using standard epidemiologic questionnaires. This will also strengthen the capacity to extract new insights from this big data.

New technologies and data sources for big data in health will enable the collection of information on health outcomes and risk factors—which has traditionally been collected via questionnaires—and aggregate these data in a rapid and cost-effective manner. Input can come from electronic health records (EHRs), innovative primary data collection tools such as new sensors of patient disease symptoms, real-time monitors of behavior, and secondary data sources such as location-linked databases of environmental and neighborhood characteristics, which we describe below.

Data on disease outcomes. Electronic health records contain a wealth of routinely collected medical information—making them the single most comprehensive source of health data—and therefore are an essential part of any large database of disease outcome data. It has been estimated that 80% of the information in an EHR is unstructured data such as scanned images or text from physician's notes, but improvements in data mining technology are making these data increasingly accessible. Recent requirements for meaningful use will require EHRs to become more effective for data input, storage, interoperability, analysis, and prediction, and future requirements could encompass more social and behavioral domains.[23]

In addition to EHRs, data on disease outcomes can also be collected from sensors and health apps. Examples include Asthmapolis, which tracks a patient's use of both rescue and controller asthma medications with an inhaler sensor, and Glooko, which monitors blood glucose levels in diabetics. While sensors are not currently available for all health outcomes, when available, they provide more detailed and timely information than is found in a typical EHR and avoids the inherent error in self-reported measures.

A final source of population-level disease outcome data is crowdsourcing.[24] For example, data on chronic diseases is collected through websites such as PatientsLikeMe and the Health Tracking Network. Wiki-type web pages are also used to manage and interpret data.[25] Crowdsourced data is especially important for monitoring the spread of infectious diseases. Google Flu Trends pioneered the analysis of daily influenza-related online search queries, which has been applied to tracking and predicting influenza outbreaks.[26] While in some cases this system has been successful in predicting outbreaks earlier than traditional surveillance systems, in other cases the predictions have been inappropriately high.[27] Another program, Flu Near You, collects weekly participant-reported flu symptoms through a website and mobile application to map influenza across the United States in real time. HealthMap collects, filters, and utilizes informal online data sources (e.g., online news aggregators, eyewitness reports, expert-curated discussions, and validated official reports) to analyze, map, and disseminate information about infectious disease outbreaks. Informal health data on social networking sites such as Facebook and Twitter are currently being studied to assess disease spread in real time (e.g., Fount.in, Sickweather). Similar crowdsourcing techniques have been leveraged for crisis mapping, in which eyewitness reports submitted via e-mail, text messages, or social media are plotted on interactive maps. These data can help target areas for emergency services and additional resources.

Data on behavioral risk factors for health. The initiation and maintenance of behavior change can be challenging, and even those interventions that succeed in experimental settings often do not scale well.[28] The first step to improving health behaviors is to monitor and measure them, and recent technological advances have provided many new ways of doing this. Devices or smartphone apps can be used to monitor health behaviors such as physical activity (e.g. FitBit; Jawbone Up, RunKeeper); diet (My Meal Mate)[29]; sleep quality (e.g. Lark); and medication adherence (MyMedSchedule).[30] These technologies enable the continuous recording of Observations of Daily Life,[31] which allows a more detailed record of behaviors and their trends over time than can be collected via questionnaires.

Placing health within context: social and environmental determinants of health. Growing understanding of the importance of environmental determinants of health has raised interest in integrating environmental and neighborhood data into health studies. The social and physical environment provides the context that can enable healthy behaviors or hinder them. For example, obesity within a person's social network has been shown to be a predictor of their own body mass index.[32] The walkability of a neighborhood can impact the amount of exercise that residents get,[33] and access to supermarkets may affect their ability to buy fresh fruits and vegetables.[34] Furthermore, the physical environment (e.g., air quality, pollution, crime, noise, public transportation access) has direct impacts on health that need to be better understood at both population and individual levels.

A large amount of environmental data is regularly collected in non-health sectors, and could be an important input into health-related big data. A few relevant examples of available data on the physical environment include weather patterns, pollution levels, allergens, land use change, forest fires, particulate matter, traffic patterns, pesticide applications, or water quality. Growing capacity for ambient environmental sensing, citizen science and the use of drones will expand access to remotely and passively captured environmental surveillance data (e.g., CitiSense).[35–37] The social and economic environment can be quantified using spatially explicit socioeconomic data, such as from the U.S. Census, American Community Survey, or publicly available crime data. And social connectedness can be assessed through online social networks.

> "WHILE SENSORS ARE NOT CURRENTLY AVAILABLE FOR ALL HEALTH OUTCOMES, WHEN AVAILABLE, THEY PROVIDE MORE DETAILED AND TIMELY INFORMATION THAN IS FOUND IN A TYPICAL EHR AND AVOIDS THE INHERENT ERROR IN SELF-REPORTED MEASURES."

Geography provides a unifying framework to integrate all of these disparate data sources. Tools such as Global Positioning Systems (GPS)[38] and Geographic Information Systems (GIS) allow multiple layers of diverse types of data to be georeferenced and layered, enabling a more complex and defined assessment of the social and environmental drivers of health.[39] The recent development of open source mapping and visualization products (e.g., Quantum GIS, OpenStreetMap, CartoDB, MapBox) is enabling the growth and application of this field in new ways.

Genomics. Genomics is an important risk factor for disease,[9] both as a direct cause of disease and as a marker of susceptibility or resistance to other risk factors. High throughput genetic information is becoming increasingly easy and inexpensive to collect. For example, one personal genetics company, 23andMe, offers DNA analysis services to consumers, as well as the option to share their personal data with 23andMe's research efforts. A research portal will share de-identified, aggregated health data with academic collaborators, which has the potential to create large databases for a fraction of the cost of traditional research programs.

Interventions to improve disease risk factors

Once research has identified the appropriate risk factor targets for a subpopulation or given person, the next step in the disease prevention process is to help that person achieve these goals. In the past this might have meant a brief word of advice from one's physician at the annual checkup to avoid smoking, exercise, and eat healthy foods. But big data offers the potential for this important advice to reach each person outside of the clinic in a personalized manner, which increases the likelihood of its impact.

Monitoring health behavior and providing real-time feedback on performance in comparison to personalized targets can help people reach their behavior goals. Ideally, behavioral data should be passively collected in order to allow continuous, long-term follow-up that does not require additional effort from the patient. This information can also be connected to a research database, as described previously, thus completing a virtuous and rapid iterative data cycle from research to intervention and back to research.

In addition to simple monitoring, a more sophisticated program would include algorithms that provide personalized feedback to assist with behavior modification at key moments of decision making (e.g., suggesting healthy recipes while the patient is shopping; encouraging exercise at the end of the workday, or giving a personalized warning about location based environmental triggers for asthma). The real-time velocity sets this application of big data apart from traditional public health uses of behavioral or health data.

Case Study 1: Big Data and Physical Activity

Physical activity, an important behavioral risk factor for chronic disease, is affected by many social and environmental factors.[40] Physical activity is most commonly assessed using questionnaires, which are of varying degrees of reliability and validity,[41] and are administered infrequently, constituting an incomplete picture of the quantity and nature of the physical activity conducted. In contrast, new devices (e.g. Fitbit, Jawbone Up, Nike FuelBand) and smartphone apps that have the potential to passively and continuously track physical activity,[42] constitute a novel source of big data for population health. If combined with health information from EHRs, these data could be used to conduct studies of how physical activity affects health, as well as how physical activity is affected by the social and environmental context. An important property of such studies is that the large study population would allow discovery of subgroups (defined perhaps by sociodemographic factors or geography) in which the key associations differ. Examples might include people with a certain metabolic profile, for whom physical activity is especially beneficial to health, or a sociodemographic group for whom improving the physical environment is insufficient to spur increased physical activity because their social context is the key limiting factor. This knowledge could be used to tailor population health interventions (i.e., precision prevention).

Big data could also directly help people improve their physical activity habits by enabling them to track and understand their own activity patterns, and support their efforts to improve them on their own. One example would be real-time reminders to increase physical activity before the end of an unusually sedentary day to avoid missing one's daily activity target. Online social networks could facilitate population health interventions to increase physical activity by linking groups in order to increase motivation. In addition to allowing novel data collection that leads to better research, big data can facilitate participant empowerment and behavior change.

> “GEOGRAPHY PROVIDES A UNIFYING FRAMEWORK TO INTEGRATE ALL OF THESE DISPARATE DATA SOURCES.”

Case Study 2: Big Data and Asthma

Over 25 million Americans suffer from asthma, and the prevalence has been increasing since the early 1980s across all age, sex, and racial groups.[43] As one of the most common chronic conditions in the United States, asthma accounted for over $56 billion in healthcare costs and lost productivity in 2007.[44] Asthma is unique in that exacerbations can be triggered by short-term variability in environmental exposures. Although many indoor and outdoor triggers of acute asthma exacerbations have been identified, few studies have been able to identify the fine-scale spatial predictors of asthma exacerbations. Public health research has relied upon data aggregated at the city, county, or even state level, and time lags between data collection and availability lead to analyses on data that are 1–2 years old. Moreover, individuals traditionally have had to manually track their events and symptoms in asthma diaries, which led to incomplete data collection and recall error.

The Asthmapolis sensor, which snaps onto asthma metered-dose inhalers, resolves many of these problems. It passively captures the time, location, and GPS coordinates of inhaler use by communicating with a smartphone. The smartphone application allows users to provide further contextual information, such as symptoms, perceived triggers, activity at time of use, and whether the inhaler was used for prophylactic purposes. These data are uploaded in real time to a remote server and detailed analytics are then provided back to the patient, their provider, and caregivers. The data also clarify relationships between controller medication adherence and asthma exacerbation events, creating a data feedback loop to improve adherence behavior. In initial studies of the use of the Asthmapolis sensor, participants experienced reduced asthma symptoms and improved control and awareness over a 4-month period.[45,46]

Beyond individual use, the analysis of the de-identified, aggregated data can be used for public health surveillance to identify local asthma hotspots across a region. When combined with contextual big data, such as environmental data sources, traffic data, or weather patterns, these spatially explicit, real-time data enable more targeted temporal and spatial analysis of environmental drivers of asthma, including methods such as small-area analyses.[39,47,48] This type of information can enable cities and regions to address asthma burden, evaluate intervention scenarios, and prevent asthma with a data-driven approach. The city of Louisville, Kentucky, has adopted this technology to address their elevated asthma burden.[49] Because the Asthmapolis system enables both individual- and population-level analyses, it achieves both quantified self and quantified community goals.

Concerns and Limitations

Obstacles exist for using big data in both research and intervention. Among the most important issues is privacy. For example, combining massive protected and personal datasets (i.e., EHRs, health behaviors, etc.) raises serious logistical and ethical challenges to maintaining privacy. As discussed in the National Research Council (NRC) report on precision medicine,[9] complex issues of consent, confidentiality, patient access, and oversight will need to be addressed in order to combine such large quantities of individual data. The current technical difficulty of combining data from different source datasets poses challenges, such as with incompatible proprietary EHRs. Because individuals change medical providers frequently and therefore have health data scattered across multiple EHR systems, EHR incompatibility will complicate the effort to assemble long-term datasets of individuals' health information. Similar difficulties will be encountered when combining environmental data that has been collected at different spatial scales, and behavioral data that has been collected on different temporal scales.

Additional unresolved challenges include the funding, administration, and accessibility of such a merged dataset of detailed health, behavioral, and environmental data. The NRC report suggested that a national database should be administered by the U.S. government, but other governance models, such as a public/academic/private-sector consortium, are also possible. The administering body will need to carefully curate both the data and the metadata, as the utility of the dataset will decrease without knowledge of a variable's source and appropriate interpretation, as well as set responsible policies for data accessibility. In spite of these logistical and ethical challenges, early examples of large-scale database linkages already exist and can yield insights into which approaches are most feasible.[14]

Beyond these challenges, discovering personalized disease risk factors from such a massive dataset will be a signal detection challenge that traditional analytic methods are ill equipped to achieve.[13] The statistical methods used will need to combine hypothesis-driven research with agnostic methods to identify susceptible subgroups. Compounding these analytic issues is a dearth of people possessing both the necessary substantive knowledge in health, as well as the big data training required to assemble, curate, and analyze such a variety of data at this scale.[50,51] We anticipate a necessary trade-off between rigor and vigor,[52] as the desire to quickly draw and apply novel insights from this data conflicts with the need to analyze the data methodically and apply the insights cautiously. This conflict may subside over time as these efforts mature and a better understanding emerges of how to balance these competing priorities.

“MONITORING HEALTH BEHAVIOR AND PROVIDING REAL-TIME FEEDBACK ON PERFORMANCE IN COMPARISON TO PERSONALIZED TARGETS CAN HELP PEOPLE REACH THEIR BEHAVIOR GOALS.”

Finally, challenges exist in the development and validation of technologies for passive, long-term monitoring of health behaviors. This validation is easier for some behaviors (e.g., physical activity) than for others (e.g., diet). An additional challenge for interventions includes optimizing the real-time algorithms that interact with users to support and encourage healthier behaviors.

Conclusions

Big data has a potentially critical role to play in preventing disease. It can both allow the discovery of new, personalized disease risk factors related to lifestyle or the environment, and also help people to successfully modify their risk behaviors. By alleviating the increasing burden of behavior-related diseases in the United States, big data could improve population health and reduce healthcare costs.

Adding urgency to this effort is the fact that the burden of such chronic conditions will be compounded by the large aging population in the United States and the increasing cost of health care.[1] Together these provide significant incentive for accelerating progress in helping people achieve improved health and well-being. Policies related to federal health care reform are placing more responsibility on healthcare providers to prevent disease, which add an additional financial incentive for this type of ambitious big data–based disease prevention effort.[49] One healthcare delivery model currently being implemented is the affordable care organization, in which healthcare providers receive capitated payments (i.e., a fixed fee per patient regardless of treatment required) to provide medical care for a defined population.[53] Under this system, the financial incentives would be aligned for healthcare providers to prevent people from becoming sick. This will necessitate discovery of new cost-effective ways to prevent disease by intervening on the determinants of health, with the goal of improving health while reducing expenditures. As we have described in this article, big data can play a key role in meeting this challenge.

Acknowledgments

M.A.B. and O.H. are Robert Wood Johnson Foundation Health & Society Scholars at the University of California–San Francisco and University of California–Berkeley. The inspiration for this paper grew from conversations in a session on big data at the Robert Wood Johnson Foundation Health and Society Scholars Annual Meeting in 2012; they thank the participants of that session. M.A.B. and O.H. thank the program for its financial support, and the faculty and scholars for their intellectual support. In particular, they thank Paula Braveman, Ralph Catalano, Ezequiel Galarce, George Kaplan, Matthew Killingsworth, Barbara Laraia, Wendy Berry Mendes, Leonard Syme, David Van Sickle, David Vlahov, and Stephen Shortell for providing valuable input.

Author Disclosure Statement

M.A.B. and O.H. are currently postdoctoral fellows employed by the University of California–San Francisco. They have conducted research projects, unrelated to the present article, using data provided by Asthmapolis, Inc. Asthmapolis had no oversight over the research process or the findings of those studies, and provided no compensation. Both M.A.B. and O.H. will accept full-time research positions with Asthmapolis after their fellowships end in September 2013.

References

1. National Research Council and Institute of Medicine. U.S. Health in International Perspective: Shorter Lives, Poorer Health. Panel on Understanding Cross-National Health Differences Among High-Income Countries. Washington, DC: The National Academies Press, 2013.
2. McGinnis J, Foege W. Actual causes of death in the United States. JAMA 1993; 270:2207–2212.
3. Mokdad A, Marks J, Stroup D, et al. Actual causes of death in the United States, 2000. JAMA 2004; 291:1238–1245.
4. National Center for Health Statistics. Health, United States, 2011: With Special Feature on Socioeconomic Status and Health. Hyattsville, MD: National Center for Health Statistics, 2012.
5. Wang YC, McPherson K, Marsh T, et al. Health and economic burden of the projected obesity trends in the USA and the UK. Lancet 2011; 378:815–825.
6. Fineberg HV. The paradox of disease prevention: celebrated in principle, resisted in practice. JAMA 2013; 310:85–90.
7. Manyika J, Chui M, Brown B, et al. Big Data: The Next Frontier for Innovation, Competition, and Productivity. San Francisco, CA: McKinsey Global Institute, 2011.
8. Chawla NV, Davis DA. Bringing big data to personalized healthcare: A patient-centered framework. J Gen Intern Med 2013, June 25. [Epub ahead of print]
9. National Research Council. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. Washington, DC: Committee on a Framework for Development a New Taxonomy of Disease, National Research Council, 2011.
10. Dyson E. The quantified community. Project Syndicate. July 23, 2012.
11. Nilsen W, Kumar S, Shar A, et al. Advancing the science of mHealth. J Health Commun 2012;17:5–10.
12. Laney D. 3D data management: Controlling data volume, velocity, and variety. META Group, 2001. Available online at blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf
13. Kumar S, Nilsen W, Pavel M, et al. Mobile health: Revolutionizing health through transdisciplinary tesearch. Computer 2012; 46:28–35.
14. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. Annu Rev Public Health 2011; 32:91–108.
15. Kaplan GA. How big is big enough for epidemiology? Epidemiology 2007;18:18–20.
16. HIMSS. A new prescription for chronic disease: Remote monitoring devices. HIMSS Analytics and Qualcomm Life, 2012.
17. Fox S. Social Life of Health Information. Washington, DC: Pew Research Center, 2011.
18. Purcell K. Half of adult cell phone owners have apps on their phones. Washington, DC: Pew Research Center, 2011.
19. eHealth. An Issue Brief on eHealth Tools and Diabetes Care for Socially Disadvantaged Populations. Washington, DC: eHealth Initiative, 2012.
20. Smith A. Smartphone ownership: 2013 update. Washington, DC: Pew Research Center, 2013.
21. The World Bank. Information and Communications for Development 2012: Maximizing Mobile. Washington, DC: World Bank, 2012.
22. World Economic Forum. Personal Data: The Emergence of a New Asset Class. Geneva, Switzerland: World Economic Forum, 2011.
23. Centers for Medicare and Medicaid Services. Meaningful Use. Centers for Medicare and Medicaid Services, 2013.
24. Swan M. Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. J Med Internet Res 2012; 14:e46.
25. Waldrop M. Big data: Wikiomics. Nature 2008; 455:22.
26. Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. Nature 2008; 457:1012–1014.
27. Butler D. When Google got flu wrong. Nature 2013; 494:155–156.
28. Rothman A. Toward a theory-based analysis of behavioral maintenance. Health Psychol 2000; 19(Suppl 1):64–69.
29. Carter MC, Burley VJ, Nykjaer C, et al. Adherence to a smartphone application for weight loss compared to website and paper diary: Pilot randomized controlled trial. J Med Internet Res 2013; 15:e32.
30. Dayer L, Heldenbrand S, Anderson P, et al. Smartphone medication adherence apps: Potential benefits to patients and providers. J Am Pharm Assoc 2013; 53:172–181.
31. Robert Wood Johnson Foundation. Tracking and Sharing observations from daily life could transform chronic care management. Project HealthDesign Blog, 2010. Available online at http://www.rwjf.org/en/about-rwjf/newsroom/newsroom-content/2010/03/tracking-and-sharing-observations-from-daily-life-could-transfor.html
32. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007; 357:370–379.
33. Cerin E, Lee KY, Barnett A, et al. Objectively-measured neighborhood environments and leisure-time physical activity in Chinese urban elders. Prev Med 2012; 56: 86–89.
34. Robinson PL, Dominguez F, Teklehaimanot S, et al. Does distance decay modelling of supermarket accessibility predict fruit and vegetable intake by individuals in a large metropolitan area? J Health Care Poor Underserved 2013; 24:172–185.
35. Boulos MNK, Resch B, Crowley DN, et al. Crowdsourcing, citizen sensing and sensor web technologies for

public and environmental health surveillance and crisis management: trends, OGC standards and application examples. Int J Health Geogr 2011; 10:67.

36. de Nazelle A, Seto E, Donaire-Gonzalez D, et al. Improving estimates of air pollution exposure through ubiquitous sensing technologies. Environ Poll 2013; 176:92–99.
37. Silvertown J. A new dawn for citizen science. Trends Ecol Evol 2009; 24:467–471.
38. Kerr J, Duncan S, Schipperjin J. Using global positioning systems in health research: A practical approach to data collection and processing. Am J Prev Med 2011; 41:532–540.
39. Elliott P, Savitz DA. Design issues in small-area studies of environment and health. Environ Health Perspect 2008; 116:1098.
40. Institute of Medicine. Accelerating Progress in Obesity Prevention: Solving the Weight of the Nation. Washington, DC: Institute of Medicine, 2012.
41. Helmerhorst HJ, Brage S, Warren J, et al. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. Int J Behav Nutr Phys Act 2012; 9:1–55.
42. Donaire-Gonzalez D, de Nazelle A, Seto E, et al. Comparison of physical activity measures using mobile phone-based CalFit and actigraph. J Med Internet Res 2013; 15:e111.
43. Akinbami OJ, Moorman JE, Bailey C, et al. Trends in Asthma Prevalence, Health Care Use, and Mortality in the United States, 2001–2010. Washington, DC: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, 2012.
44. Barnett SBL, Nurmagambetov TA. Costs of asthma in the United States: 2002–2007. J Allergy Clin Immunol 2011; 127:145–152.
45. Van Sickle D, Maenner M, Barrett M, et al. Monitoring and improving compliance and asthma control: mapping inhaler use for feedback to patients, physicians and payers. Resp Drug Delivery Europe 2013; 1:119–130.
46. Van Sickle D, Magzamen S, Truelove S, et al. Remote monitoring of inhaled bronchodilator use and weekly feedback about asthma management: An open-group, short-term pilot study of the impact on asthma control. PloS One 2013; 8:e55335.
47. MacDonald C. New technology helps doctors link a patient's location to illness and treatment. The Washington Post, February 4, 2013.
48. Weiss KB, Wagener DK. Geographic variations in U.S. asthma mortality: Small-area analyses of excess mortality, 1981–1985. Am J Epidemiol 1990; 132:107–115.
49. MacDonald C. Using big data to improve health: Geomedicine combines pollution and health data to better inform patients, doctors and researchers. The Environmental Magazine, November 1, 2012.
50. Dumbill E, Liddy ED, Stanton J, et al. Educating the next generation of data scientists. Big Data 2013; 1:21–27.
51. Davenport TH, Patil D. Data scientist: The sexiest job of the 21st century. Harv Bus Rev 2012; 90:70–77.
52. Adler N, Bush NR, Pantell MS. Rigor, vigor, and the study of health disparities. Proc Natl Acad Sci 2012; 109(Suppl 2):17154–17159.
53. DeVore S, Champion RW. Driving population health through accountable care organizations. Health Aff 2011; 30:41–50.

Corresponding author:
Address correspondence to:

Olivier Humblet
Center for Health and Community
University of California–San Francisco
3333 California, Suite 465
San Francisco, CA, 94143

E-mail: HumbletO@chc.ucsf.edu