Balancing Big Data and the Right to Health: Strategies for Maximising Ethical and Sustainable Impact

Lucy Fagan, Global Health Next Generation Network*

Introduction

In the field of health care, the question of who – frontline doctors or data – should decide which patients receive what kinds of care has been the subject of many debates in recent years (Pulitzer Center, 2016). Meanwhile, the advent of "big data" has transformed the volume, velocity, and character of the information that we are able to procure regarding virtually every aspect of human life (Boyd & Crawford, 2012). These large datasets, along with dramatic advances in computational power, have allowed us the capacity to collect and analyse data on an unprecedented scale, linking and comparing vast quantities of previously incomparable data that can now be used to provide evidence to inform policies and practices (Lazer, 2009; Boyd & Crawford, 2012).

Big data provides an opportunity to reduce cost and time of monitoring and evaluation by transforming the ways in which we assess policies and programs (Raghupathi & Raghupathi, 2014; Groves, 2013). It also offers the opportunity to create efficiencies and identify trends and opportunities that would otherwise go unnoticed (Murdock & Detsky, 2013). This data, which is drawn from our mobile phones, web searches, health records, and beyond, seeps into nearly every facet of our digital lives (Klauser & Albrechslund, 2013). But not all persons live digital lives, and for these individuals, the era of big data can mean their fall into obscurity. As the world advances scientifically and technologically, countries face a difficult challenge in thinking beyond automated data and cost-effectiveness findings to ensure that no one is left behind.

Arguments have been made that big data should not be the sole basis for priority-setting and decision-making. Context matters too, but this may be lost in datasets where the value of life has been reduced to statistics and numerical values. Using big data requires critical thought on the quality of our sources, the assumptions inherent in the questions we ask and the conclusions we draw, and how the available information fits within the goals and objectives of the program or policy in question (Boyd & Crawford, 2012). Perhaps more importantly than what the data captures is what it doesn't capture: groups of people who are excluded from collection mechanisms because of circumstances of poverty, geography, or culture.

To maximise the positive impact of big data, it is important to know both the potential and the bounds of these datasets, the methods used to gather and analyse them, and the ways in which they are interpreted. By prioritising the universal right to health, leveraging the expertise of multi-sectoral teams through team science and translational science, and using data wisely, it is possible to advance our capacity to make data-driven decisions in a way that is equitable, ethical, just, and sustainable.

Limitations of Big Data for Health Policymaking

Using big data has many advantages: these datasets are often integrated into existing systems, thus reducing the burden of collection, and the algorithm-based functioning of many analytic processes allows for vast quantities of data to be gathered and reported in real-time. The advantages of this type of information are not to be understated.

However, this readily accessible data may not always capture the variables that we want or need to know in order to draw accurate and representative conclusions, and quantity of data does not necessarily equate to quality of data (Boyd & Crawford, 2012). Automatically aggregated data or poorly constructed analytic frames may be susceptible to biases, weaknesses, and inferences (Lazer et al., 2014).

*The views and opinions expressed are the author's and do not represent those of the Secretariat of the United Nations. Online publication or dissemination does not imply endorsement by the United Nations.

More importantly, we must consider what - and whom - is obscured by our usage of these large-scale datasets. Country-level data, for example, may obscure dramatic regional or socioeconomic variations; similarly, if analytic frames are not sufficiently calibrated to account for important subgroups, big data may miss subtle trends. Certain segments of the population may not contribute to the data pool, including those without access to the technology used to collect data. Others, such as minority populations or those experiencing rare health conditions, may not register as significant in the scope of all of the data collected, and thus, may not garner the appropriate attention. Though they are not unique to big data, these absences and failures in representation pose serious concerns.

Ethical Dilemmas at the Nexus of Big and Small Data

Consent, privacy, and security are among the primary issues with implementation of these data collection and analytic schemes; these concerns extend to the implementation of predictive analytic models (Cohen et al, 2014). Individuals may not always comprehend the extent to which they are providing data, or anticipate how that data will be used, and health information is a particularly sensitive subject (Cate & Mayer-Shönberger, 2014). As these datasets become increasingly detailed and connected, and thus data points become increasingly personally identifiable, security becomes an even larger concern. Information in these datasets has high black-market value. If data cannot be collected with truly informed consent, and protected sufficiently, it is irresponsible and unethical to gather and use this data.

These datasets are frequently used to prioritise issues and make programming decisions; cost-effectiveness is a major aspect of these conversations and the subject of many discussions on ethics (Pinkerton et al., 2002). The means by which we determine whether or not a policy, program, or procedure is cost-effective are complicated, and the determinations we make are ultimately heavily entrenched within our own value systems. When a program is more effective, but more expensive (or alternatively, less effective but less expensive), a value judgement must be made on whether or not to allocate resources to that program or to an alternative. These evaluations come with

their own set of ethical quandaries, and the fact that an initiative is determined to be cost-effective does not mean that funds are available to operationalise it. Ultimately, cost-effectiveness may not always be the best measure for determination of whether a program or policy should be implemented; there may be times when a cost-ineffective measure is critical for development of infrastructure or simply for providing high-quality care to patients.

However, these value judgements, which are made on a macro-scale, do not always translate well to health care providers making decisions about care on the ground. These health care providers have a professional mandate to heal and provide healthcare to care-seeking patients, a value system which may lead to actions which are inconsistent with what is indicated based on a big data driven decision-making strategy.

Recommendations

Foremost, keep people at the centre of health policymaking and decisions: **health is a UN-affirmed human right, and should be prioritised as such** (United Nations, 1948). While strategic resource allocation is important, policymaking should focus more on how to provide health care than for which types of health issues to provide care. There is a fundamental difference between choosing the most cost-effective strategy for targeting a particular health problem, and choosing which health problem is most cost-effective to target. Ultimately, all humans have a right to health irrespective of their socioeconomic status, their geographic location, or the cost-effectiveness of their condition. In denying the universal right to human health, we purposefully exclude certain groups of people from accessing the care they need.

Second, leverage multi-sectoral expertise: **collaborative expertise is critical** in designing comprehensive data collection and analytic systems, in ensuring equitable analytic frames, and developing actionable and sustainable evidence-based policy. *Team science*, in which teams are comprised of scientists from across a variety of fields, will be especially critical in ensuring that big data is collected, analysed, and interpreted in a way which is equitable, particularly because scientists and program managers from a diverse array of backgrounds have a stake in the development and

use of these data systems. *Translational science,* which focuses on strategically applying innovative research to meaningful and actionable policy, is likewise an essential area of focus for both scientists and policymakers. Training for young scientists should emphasise not only cross-disciplinary engagement but also tools for the application and dissemination of research to policymakers and program managers.

Finally, be attentive: **big data needs to be gathered, analysed, and used wisely**. These datasets should be used to supplement existing data collection mechanisms rather than to substitute for them, not only as a means of maintaining quality of data but also as a mechanism for inclusion of all populations. Strengthening measures to protect patient consent and privacy are critical. To manage the ethical dilemmas which permeate the process of health care decision making, data collection, and analytics, as well as policy development, rigorous ethics training will be valuable not only for young scientists but also for seasoned professionals and policymakers. Application of data-drawn conclusions should be done with an awareness of the assumptions inherent in the mechanisms of collection, the analytic methods, and the inherent biases in conclusions drawn.

Big data is transforming the ways in which we engage with human health; in order to ensure that it is done in a way which is equitable and sustainable, we must be conscious of the way in which we gather, analyse, and utilise the information we glean.

References

Boyd & Crawford (2012) Critical Questions for Big Data (http://www.tandfonline.com/doi/pdf/10.1080/1369118X.2012.678878)

Cate & Mayer-Shönberger (2014). Notice and Consent in a World of Big Data (http://idpl.oxfordjournals.org/content/3/2/67.abstract)

Cohen et al. (2014) The Legal and Ethical Concerns that Arise from Using Complex Predictive Analytics in Health Care (http://content.healthaffairs.org/content/33/7/1139.abstract)

Groves et al. (2013) The 'Big Data' Revolution in Healthcare: Accelerating Value and Innovation (http://www.pharmatalents.es/assets/files/Big_Data_Revolution.pdf)

Klauser & Albrechtslund (2014) From Self-Tracking to Smart Urban Infrastructures: Towards an Interdisciplinary Research Agenda on Big Data (http://library.queensu.ca/ojs/index.php/surveillance-and-society/article/view/infrastructures/infrastructure)

Lazer et al. (2014) The Parable of Google Flu: Traps in Big Data Analysis (http://gking.harvard.edu/files/gking/files/0314policyforumff.pdf?m=1394735706)

Lazer et al. (2009) Computational Social Science (http://science.sciencemag.org/content/323/5915/721)

Murdoch & Detsky (2013) The Inevitable Application of Big Data to Health Care (http://jama.jamanetwork.com/article.aspx?articleid=1674245)

Pinkerton et al. (2002) Ethical Issues in Cost-Effectiveness Analysis (https://www.uic.edu/sph/prepare/courses/chsc433/pinkerton.pdf)

Pulitzer Center (2016) The Life Equation (http://pulitzercenter.org/project/life-equation)

Raghupathi & Raghupathi (2014) Big Data Analytics in Healthcare: Promise and Potential (http://www.hissjournal.com/content/2/1/3)

United Nations (1948) Universal Declaration on Human Rights, Article 25 (http://www.un.org/en/universal-declaration-human-rights/)