21

Medical Privacy and Big Data

A Further Reason in Favour of Public Universal Health-Care Coverage

*Carissa Véliz**

Most people are completely oblivious to the danger that their medical data undergoes as soon as it goes out into the burgeoning world of big data. Medical data is financially valuable, and your sensitive data may be shared or sold by doctors, hospitals, clinical laboratories, and pharmacies—without your knowledge or consent.[1] Medical data can also be found in your browsing history, the smartphone applications you use, data from wearables, your shopping list, and more. At best, data about your health might end up in the hands of researchers on whose goodwill we depend to avoid abuses of power.[2] Very likely, it will end up with data brokers who might sell it to a future employer, or an insurance company, or the government. At worst, your medical data may end up in the hands of criminals eager to commit extortion or identity theft. In addition to data harms related to exposure and discrimination, the collection of sensitive data by powerful corporations risks the creation of data monopolies that can dominate and condition access to health care.

This chapter aims to explore the challenge that big data brings to medical privacy. Section I offers a brief overview of the role of privacy in medical settings. I define privacy as having one's personal information and one's personal sensorial space (what I call *autotopos*) unaccessed. Section II discusses how the challenge of big data differs from other risks to medical privacy. Section III is about what can be done to minimize those risks. I argue that the most effective way of protecting people from suffering unfair medical consequences is by having a public universal health-care system in which coverage is not influenced by personal data (eg genetic predisposition, behavioural data, etc).

I. Medical Privacy

Privacy consists in having one's personal information and one's personal sensorial space (what I call *autotopos*[3]) unaccessed.[4] Personal information is the kind of information that

* This chapter has been written with the support of the Wellcome Trust (grant number WT104848/Z/14/Z).

[1] A Tanner, *Our Bodies, Our Data: How Companies Make Billions Selling Our Medical Records* (Beacon Press, 2017), 1.

[2] Ibid.

[3] From the Greek *auto* (self) and *topos* (place or space). My thanks to Roger Crisp for suggesting the term.

[4] The adjective 'unaccessed' is not found in dictionaries, but there is no suitable existing term to convey in one word the property of not having been accessed. 'Inaccessible' denotes the property of not being able to be accessed, which is different from being accessible yet not actually accessed. Analogous differentiations exist in English that use the same prefixes (eg indisputable/undisputed, inalterable/unaltered, etc).

Carissa Véliz, *Medical Privacy and Big Data* In. Edited by: Andelka M Phillips, Thana C de Campos, and Jonathan Herring, Oxford University Press (2019).
DOI: 10.1093/oso/9780198796558.003.0022

someone in a certain society would not want to share with just anyone, or information an individual is particularly sensitive about. The autotopos is the kind of sensorial space that people in a certain society would not want just anyone, other than themselves (and perhaps a very limited number of other people chosen by them), to access. One's autotopos is accessed when someone sensorially enters a culturally established personal zone of one. That is, when another (through direct or indirect perception, such as cameras and microphones) sees, hears, smells, or touches one in a zone in which there are cultural expectations to be free from the eyes, ears, touch, and presence of others (eg in the toilet). One's autotopos is also accessed when one is witnessed engaging in some activity or being the subject of some event that typically evokes the desire to have no witnesses or very few chosen witnesses (eg being naked).[5]

Medical privacy refers to having personal information about one's health status unaccessed and having one's autotopos unaccessed in the context of medical settings. Some scholars, such as Anita Allen, include within medical privacy things like 'associational privacy (eg intimate sharing of death, illness, and recovery); proprietary privacy (eg self-ownership and control over personal identifiers, genetic data, and biospecimens); and decisional privacy (eg autonomy and choice in medical decision making).'[6] This over-inclusion is mistaken. What Allen calls 'associational privacy' can be explained through a combination of informational and sensorial access: we do not want non-intimate people to *see* us when we are ill or dying; we do not want them to *know* about our illnesses. Similarly, with so-called 'proprietary privacy', if we worry about the information that genetic data carries and the possible consequences of that information getting shared inappropriately, then it is a matter of privacy, but one that is covered by informational privacy. If the worry is rather financial (ie about who should profit from patients' genetic data), then it is a proprietary issue, not a privacy one. Finally, what Allen calls 'decisional privacy' is rather a matter having to do with the interest we have in being allowed to do as we please without interference.[7] In other words, it is a matter of freedom and autonomy, not privacy.

One might wonder what makes the two species—informational and sensorial access—part of the genus of privacy. The unity of the category of privacy is founded on the notion of being personally unaccessed and the kinds of interests we have in not being accessed by others. Privacy protects us from: (1) certain kinds of harms that may come about as a result of other people having access to our personal life (eg discrimination, identity theft, etc); (2) the demands of sociality; (3) being judged and possibly ridiculed by others (and thus from self-conscious negative emotions such as shame and embarrassment); and (4) the discomfort of being watched, heard, and so on.[8]

The medical context is an important home for privacy. Visits to the doctor's office or the hospital create moments of great vulnerability for patients and their families. People typically do not like just anyone knowing about their diseases (particularly when it comes to certain kinds of diseases that carry more stigma with them), and we do not like being seen at our worst—when we are sick, worried, and stripped down of makeup and our everyday attire. Yet, privacy losses are inevitable in medical contexts. For patients to get adequate

[5] C Véliz, 'On Privacy' (DPhil thesis, University of Oxford, 2017), ch 4.

[6] A Allen, 'Privacy and Medicine' in EN Zalta (ed), *The Stanford Encyclopedia of Philosophy* (winter 2016 edn), https://plato.stanford.edu/archives/win2016/entries/privacy-medicine/ accessed 15 July 2019.

[7] WA Parent, 'Privacy, Morality, and the Law' (1983) 12 Philosophy and Public Affairs 269.

[8] Véliz (n 5).

care, they have little choice but to surrender both sensitive information about themselves and access to their bodies to health-care professionals. Likewise, for subjects to participate in health research, they must give up some degree of informational and sensorial privacy. It is the duty of health-care professionals and researchers, however, to minimize privacy losses, to show consideration towards patients and subjects, and to avoid any unnecessary exposure.

Thus, health-care professionals take care to protect patients' autotopos. Sensorial privacy is protected through bedside manners (eg looking away while the patient undresses), and other practices designed to limit exposure to patients' autotopos (eg the use of hospital gowns, curtains, etc).

Similarly, confidentiality protects informational privacy. Confidentiality refers to the moral duties of non-disclosure of information shared in the context of a fiduciary, contractual, or professional relationship, such as that of the lawyer–client relationship or the doctor–patient one. The Hippocratic Oath, which set the historical foundations of medical ethics, already took into account privacy, and included a vow not to speak of what is seen and heard in the course of treatment.[9] In addition to being a means of respecting a patient's right to privacy, confidentiality protects a patient's autonomy by sheltering them from external interference and possible criticism about their decisions.[10] Confidentiality also protects patients from possible stigmatization and discrimination (eg at work, or within the family). It makes it more likely that patients will seek medical care (thus keeping medical costs lower through routine check-ups that allow prompt medical attention), be honest with doctors (making medical care more accurate), and be more willing to participate in health research. [11]

In the context of medical research, ethics committees and institutional review boards make sure research subjects' privacy is adequately protected. Confidentiality requires that research subjects not be identifiable. To fulfil that objective, researchers anonymize data through aggregating it and using 'identifiers' (as opposed to names).[12] Further protection of privacy is traditionally afforded to research subjects through the use of informed consent. In order to give consent, participants must know what data will be collected about them, and how it will be managed. Subjects should also be informed of their right to withdraw from research at any time, and they should be told what would happen to their data in that scenario (ideally, data should be destroyed).

Anonymization and informed consent have never been perfect tools for protecting privacy. Genetic data, for example, resists anonymization because it is unique to each individual. Similarly, in some cases, it is impossible to extract valid informed consent (eg if a research subject is incapacitated). Notwithstanding their very significant limitations, however, it is widely accepted that they are usually good enough tools for protecting privacy, particularly given that they can be complemented with other good practices to compensate for their shortcomings. For instance, a study using genetic data can be kept in a

[9] 'Oath of Hippocrates' in WT Reich (ed), *Encyclopedia of Bioethics* (Macmillan, 1995), 2632.

[10] Allen (n 6).

[11] For more on the legal nature of confidentiality, see J Herring, *Medical Law and Ethics* (6th edn, Oxford University Press, 2016), 224.

[12] Cited by Anita Allen (n 6); Institute of Medicine, Committee on the Role of Institutional Review Boards in Health Services Research Data Privacy Protection, *Protecting Data Privacy in Health Services Research* (National Academies Press, 2000), https://www.nap.edu/read/9952/chapter/1 accessed 15 July 2019.

closed environment (eg using computers that are not connected to the internet, and in a locked room to which only authorized researchers have access, etc) to make sure that data is not linked to other data that could expose individuals. Similarly, in cases in which informed consent from a patient is not possible, consent from family members can act as a proxy. As it will become apparent in the next section, big data complicates the use of both anonymization and informed consent to the point of risking making them null.

After this brief overview of the role of privacy in medical settings, I will go on to detail the opportunities and privacy perils that big data presents in medical settings.

II. The Challenge of Big Data for Medical Privacy

The arrival of big data promises to revolutionize medicine. With technological innovation, however, often come new ethical challenges. As we find ourselves in novel situations, the need to foresee possible risks and benefits becomes crucial to make the most of available technologies, while avoiding as many negative consequences as possible.

In what follows, I will cover some of the medical advantages that big data may offer, before going on to explore possible privacy pitfalls and suggestions to minimize risks.

A. Big Data in Medicine

The most influential definition of big data outlines it in terms of three dimensions: volume (scale), velocity, and variety.[13,14] Big data's value rests in its capacity to combine large amounts of information, faster than ever, and from a variety of sources, into a single dataset that can allow the identification of correlations that would otherwise remain undetected.[15]

Data aggregated from sources outside of traditional medical settings can be very valuable for medical purposes. At Johns Hopkins, for example, researchers could predict the location and time of a flu outbreak based on tweets.[16] At Boston Children's Hospital, researchers could predict, track, and map obesity rates in a neighbourhood using Facebook 'likes'.[17]

It is widely believed that big data holds incredible potential to advance the diagnosis, treatment, and prevention of diseases through resolving some current problems in medicine. Randomized controlled trials (RCTs) have been crucial in the development of rigorous

[13] D Laney, '3D Data Management: Controlling Data Volume, Velocity and Variety' (2001) 6 META Group Research Note.

[14] Laney's conceptualization has become the classic definition. Subsequent approaches have added other elements to big data. See eg M Ali-ud-din Khan, MF Uddin, and N Gupta, 'Seven V's of Big Data: Understanding Big Data to extract Value' (Proceedings of Zone 1 Conference of the American Society for Engineering Education, Bridgeport, CT, April 2014), doi: 10.1109/ASEEZone1.2014.6820689. In their paper, Khan et al add validity, volatility, veracity, and value to elements of big data. Most of these components, however, seem more normative than descriptive. Sometimes, data used in big data analyses have little veracity or validity, leading to errors. While it is important to emphasize the desirability of features such as veracity and validity to create value, they do not seem intrinsic to a definition of big data. Inaccurate data can still form part of big data.

[15] JH Thorpe and EA Gray, 'Big Data and Public Health: Navigating Privacy Laws to Maximize Potential' (2015) 130(2) Public Health Reports 171.

[16] DA Broniatowski, MJ Paul, and M Dredze, 'National and Local Influenza Surveillance through Twitter: An Analysis of the 2012–2013 Influenza Epidemic' (2013) 8(12) PLoS One e83672.

[17] R Chunara, L Bouton, JW Ayers, et al, 'Assessing the Online Social Environment for Surveillance of Obesity Prevalence' (2013) 8 PLoS One e61373.

medicine by providing evidence on the efficacy and safety of drugs and interventions of various kinds. However, RCTs are expensive; their findings are both too broad (given problems of statistical sampling, a treatment found to be beneficial in a trial may not be beneficial for any given individual) and too narrow (trial population and setting may not be representative of the general practice); the randomization of patients may be ethically dubious, as some patients will receive better treatment than others, whether the better treatment is the existing care or the intervention being tested; and there are usually long delays before RCT results can translate into common practice. Big data aspires to solve these issues by having access to inexpensive data—generated as a by-product of patient care and people's day-to-day lives—that is specific to individuals and that assembles information from large groups of people, therefore being adequately narrow and broad. Big data research also avoids the ethical dubiousness of randomization, and treatment could potentially be much more immediately available.[18]

Big data may thus greatly advance personalized medicine. At least for some kinds of diseases, such as cancer, treatments are sometimes only effective in a small number of patients. There are many different varieties of cancer, and people with different genomes react differently to drugs. Big data holds the promise of being able to develop precision medicine on the basis of genomic profiles.[19] Rather than prescribe a drug tested on a small sample of the population in the hopes that a patient will react the way most people did in the sample, analyses based on big data will not have to hope for the best. Instead of working from a mere sample of the population, all subjects can be profiled individually, and the drug that has been shown to work better for people closest to a patient's profile can be prescribed.

Finally, big data may uncover hitherto unsuspected correlations by taking into account data about people's living environments, their responsibility for their health (exercise, eating, and drinking habits, etc), social relations, and more.

Although the promise of big data in medicine is great, it is worth taking into account that, so far, it is mainly just a promise, and one that has been there for many years. For all their limits, RCTs are still the best source of medical evidence we have, and there is a chance big data may never live up to its many expectations. As Sir Richard Peto, Professor of Medical Statistics and Epidemiology at the University of Oxford, points out: 'You need large-scale, randomized evidence to answer a lot of questions, and I think the claim that database analysis will do so isn't justified.'[20]

B. Privacy and Other Derivative Risks

All the possible medical benefits of big data, assuming they will come, will arise as a result of having more data about patients. With more data, however, come more privacy risks. Any time data is collected, there is a risk that information may be abused or may end up in the wrong hands. In the past, however, risks were partly minimized by technological limits. When health records or data from studies were collected and stored in paper in hospitals

[18] DC Angus, 'Fusing Randomized Trials with Big Data: The Key to Self-learning Health Care Systems?' (2015) 314 Journal of the American Medical Association 767.

[19] J Andreu-Perez, CC Poon, RD Merrifield, et al, 'Big Data for Health' (2015) 19(4) IEEE Journal of Biomedical Health and Informatics 1193, 1197.

[20] Quoted by Tanner (n 1) 164.

and doctors' offices, they were usually kept under lock and key in a cabinet, and the possibility of it reaching a great number of people was highly unlikely. As soon as health records become electronic, and are stored in computers with internet access, the risk of data breaches, leaks, and misuses increases. Any data kept on a device connected to the internet is potentially hackable. And any hacked data can end up being exposed and possibly sold on the dark web. Data breaches are common in medical settings. Just in 2015, a particularly bad year, over 112 million health records were breached in the United States alone.[21] While the number of people (health records) affected was lower in 2017, the number of healthcare data security incidents was higher than in past years, and seems to be on the rise, which suggests that patients' health records are increasingly at risk.[22]

In addition to the risk of having Electronic Health Records (EHRs), privacy risks are exacerbated by big data through the amount of data collected and the sources it comes from. Medical data not only comes from traditional sources like EHRs and genetic and microbiomic sequencing data, but also from smartphone applications, data brokers, social networks, internet searches, environmental data (eg air quality data), ambient sensors, wearables, etc.[23] Health data reaches beyond the doctor's office and the hospital into everyday life. Big data will be able to correlate relationships between health and buying habits, movement tracking, sleeping habits, social relations, and more. The objective is to be able to answer questions such as: Do Facebook friends influence one's life choices? Is using a bicycle to get to work healthier than walking? Is the air in some cities so polluted that it can affect life expectancy?[24]

Thus, data that may be used for medical (and other) purposes can include data that subjects voluntarily give up, data inferred from aggregated data, data that individuals may have no knowledge is being collected, and data that could potentially be collected against the wishes of data subjects (eg browsing history). All this information can be stored indefinitely, which, in time, allows data to be used for purposes other than that for which it was collected.[25] Some of these uses can hurt data subjects.

Inappropriate uses of medical data abound and will probably increase as the data economy expands. Private corporations' use of medical data is an important concern when companies get involved with health care. For example, Google's DeepMind has been involved with the NHS in the United Kingdom to develop a clinical application for kidney injury. Millions of medical records were shared with the company without informing patients or asking for their consent. Many months later, the United Kingdom's Information Commissioner's Office ruled it unlawful.[26] Even though DeepMind had made public assurances that the medical data would not be linked to Google accounts, products, or services, there was never any legal guarantee that they would keep their word. The company has

[21] D Munro, 'Data Breaches in Healthcare Totaled over 112 Million Records in 2015', *Forbes* (31 December 2015), https://www.forbes.com/sites/danmunro/2015/12/31/data-breaches-in-healthcare-total-over-112-million-records-in-2015/#5118fabc7b07 accessed 15 July 2019.

[22] HIPAA Journal post, 'Largest Healthcare Data Breaches of 2017' (4 January 2018), https://www.hipaajournal.com/largest-healthcare-data-breaches-2017/ accessed 15 July 2019.

[23] Andreu-Perez et al (n 19) 1194.

[24] GM Weber, KD Mandl, and IS Kohane, 'Finding the Missing Link for Big Biomedical Data' (2014) 311 Journal of the American Medical Association 2479.

[25] Thorpe and Gray (n 15).

[26] J Powles, 'Why Are We Giving Away Our Most Sensitive Health Data to Google?', *The Guardian* (5 July 2017), https://www.theguardian.com/commentisfree/2017/jul/05/sensitive-health-information-deepmind-google accessed 15 July 2019.

recently been accused of breaking its promise after it announced that Google would absorb DeepMind's medical division.[27] At the moment, because it is not in the position of a health-care professional, Google does not owe duties of confidentiality to patients. People's medical data could be used for all sorts of purposes: to be sold to third parties, for targeted marketing, for personalized pricing (ie selling certain products at a higher price to people who want or need them the most), or for developing Artificial Intelligence tools that can be applied in unknown ways in the future. It has been suggested that DeepMind's interest in helping develop this application relied from the start on acquiring data for its machine-learning commercial research projects.[28] A corporation may use sensitive data surrendered in the belief that it might help patients in questionable ways.

A related significant concern that is not directly about privacy, but rather a derivative worry about the power that comes with access to private data, is that big corporations that have more financial and technological capabilities than governments will monopolize access to health care. Dominance in data collection leads to monopolies, and dominance in the collection of sensitive data can lead to particularly problematic monopolies. Google, for example, could hold a monopolistic position of power over health analytics on account of being the institution holding more data on individuals around the world. They may develop the most advanced algorithms for medical diagnosis with that data, and other possible competitors who do not have as much data will not stand a chance against the titan. As things stand, DeepMind is getting patients' data for free, and yet it could potentially retain full power over the knowledge and algorithms that it develops from its collaboration with the NHS—as well as the profits.[29] If we allow corporations to hold power over data and access to health care, prices could soar, and medical attention could be given to individuals only under unacceptable data deals (eg in addition to payment, forcing patients to give up all of their personal data in exchange for medical care).

Other privacy-related risks of big data include people being discriminated against at their workplace on account of their medical history (eg for suffering from a certain disease, or because they are pregnant). Insurance companies could also take advantage of medically relevant information to charge some people more than others (eg penalize those who do not exercise enough, or worse, those who have genes that are deemed risky). Pharmaceutical companies could engage in price discrimination by identifying people who desperately need a medicine that can only be bought from them and charge more for it. Criminals can also extort patients, threatening to expose sensitive images or information about them if they do not give up money. In 2017, a criminal group gained access to data from a cosmetic surgery clinic and extorted patients, asking for a bitcoin ransom. Hackers ended up publishing more than 25,000 private photos, including nude ones, and personal data including passport scans and national insurance numbers.[30] Another common criminal act

[27] M Murphy, 'Privacy Concerns as Google Absorbs DeepMind's Health Division', *The Telegraph* (13 November 2018), https://www.telegraph.co.uk/technology/2018/11/13/privacy-concerns-google-absorbs-deepminds-health-division/ accessed 15 July 2019.

[28] J Powles and H Hodson, 'Google DeepMind and Healthcare in an Age of Algorithms' (2017) 7(4) Health and Technology 351, https://link.springer.com/article/10.1007/s12553-017-0179-1 accessed 15 July 2019.

[29] Ibid.

[30] A Hern, 'Hackers Publish Private Photos from Cosmetic Surgery Clinic', *The Guardian* (31 May 2017), https://www.theguardian.com/technology/2017/may/31/hackers-publish-private-photos-cosmetic-surgery-clinic-bitcoin-ransom-payments accessed 15 July 2019.

is medical identity theft, committed by uninsured individuals who need medical care and steal another person's identity to get it.[31]

In a nutshell, the challenge for big data in medicine is to develop personalized medicine with the knowledge and consent of data subjects, protecting people's privacy, and minimizing risks that stem from the collection and use of sensitive data. In what follows, I explain why anonymization and consent are necessary but insufficient tools to meet this challenge.

C. The Limits of Anonymization and Consent in the Context of Big Data

As mentioned in section I above, medical ethics has relied on anonymization and informed consent to protect people's privacy. Big data, however, makes both of these tools difficult, if not impossible, to implement. Let us consider anonymization first.

Anonymization has always been a challenge. Studies that involve photographs, for example, or genetic data, can be hard or impossible to anonymize.[32] But, for the most part, other kinds of data could be more easily anonymized by stripping away information that could identify individuals. The use of big data, however, makes anonymization a near impossible feat, because the more data points we have about individuals, the easier it is to identify them. [33] There is usually only one individual of your height who works and lives where you do, for example. It only takes two or three data points to identify anyone.[34] Given the increasing amount of publicly available data that we have on people, re-identification will continue to get easier.[35] Furthermore, for data to reveal insights, it must be linked to the correct person to ensure appropriate diagnosis and treatment.[36] Therefore, every person must be uniquely tagged with a medical identifier, making it all the more easy to re-identify individuals.

Even if it were possible to anonymize data in a secure way, the analysis of aggregated data can also result in group-level harms such as stigmatization and discrimination. Such harms can impact everyone in a given group and not only the people who might have consented to research.[37]

Such are the limits of anonymization in the context of big data. Let us now turn to consent. Big data is designed to reveal unforeseen correlations, which implies that there will be a significant degree of uncertainty about future findings. Even if data subjects were to agree to give up their data, consent cannot be fully informed because subjects cannot be told about future uses and consequences of their data, as not even researchers can know what kinds of correlations may be unveiled, and they often cannot guarantee how this data will be used.[38]

[31] Tanner (n 1) 98.

[32] Nuffield Council on Bioethics, *The Collection, Linking and Use of Data in Biomedical Research and Health Care: Ethical Issues* (Nuffield Council on Bioethics, February 2015).

[33] Weber et al (n 24).

[34] YA de Montjoye, CA Hidalgo, M Verleysen, et al, 'Unique in the Crowd: The Privacy Bounds of Human Mobility' (2013) 3 Scientific Reports 1376; YA de Montjoye, L Radaelli, VK Singh, et al, 'Identity and Privacy. Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata' (2015) 347 Science 536.

[35] Tanner (n 1) 91.

[36] Andreu-Perez et al (n 19) 1204.

[37] BD Mittelstadt and L Floridi, 'The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts' (2016) 22(2) Science and Engineering Ethics 303.

[38] Ibid.

One might be tempted to think that consent is unnecessary because the risks involved in big data research are minimal. However, given the uncertainty over what kinds of information may be revealed in the future on the basis of collected data, and given the possibility of leaks and hacks, it is hard to make a case that the risks are indeed minimal.[39]

Anonymization and informed consent, then, cannot protect medical privacy in the digital age, and they do nothing to avoid concerns about power that result from the collection of sensitive data. If the traditional tools of medicine are insufficient to protect patients in big data contexts, how can risks be minimized? I turn to this question next.

III. Minimizing Risks

Part of what made anonymization and consent appropriate tools in the past were complementary practices that could strengthen protection, such as keeping paper files in a locked cabinet to which only approved researchers had access. In what follows are some complementary practices that can help minimize risks in the era of big data.

Data practices must be better regulated to protect data subjects.[40] Inappropriate uses of data, including the re-identification of individuals in anonymized databases and discrimination, should be made illegal. Similarly, we should make it illegal to link health information from research databases to other data resources if data subjects have not given their explicit consent.[41] If risks will be imposed on data subjects (eg by inferring sensitive information from non-sensitive information, or by storing data in a way that could allow for identification and misuse), medical data should not be used from people who have not agreed to be research subjects, as when data is analysed from social media or open web forums, or when doctors, hospitals, or pharmacies share medical data with data analysis companies without patients' knowledge. Policing such behaviour, however, can be quite a challenge, and even with strong regulations in place, risks from privacy losses do not disappear. For this reason, limits to how data is used are not enough, and further limits on the *access* to data should be put in place.

Consent has always been a valuable guardian of access to data, but its practice must be updated to fit big data contexts. One possibility is to put the onus on patients—to give them control of their data and make them responsible for it. Even acknowledging that the kind of consent being secured is limited and not fully informed due to the uncertainty surrounding big data, individuals could be asked for consent for each use of their data. This approach is incredibly burdensome, however, both for institutions and individuals. . Making individuals responsible for their health data will likely result in them being overwhelmed with consent requests, which may lead to them oversharing and regretting it when it is too late to

[39] Ibid.

[40] The General Data Protection Regulation (GDPR) is a step in the right direction, but it might not be enough to protect data subjects' medical data; it is too early to tell, as much depends on how the Regulation will be interpreted and enforced. The recommendations here go beyond the GDPR and are specific to medical data. See Council Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ 2 119/1, http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1525272154893&uri=CELEX:32016R0679 accessed 15 July 2019.

[41] IS Kohane and RB Altman, 'Health-Information Altruists—A Potentially Critical Resource' (2005) 353(19) New England Journal of Medicine 2074.

recall their data. A more promising proposal is 'tiered' consent, in which data subjects can choose specific future uses of their data (eg someone can decide their data can be used only for cancer research).[42] Along these lines, researchers in the United Kingdom have developed an interface that allows patients to give 'dynamic consent' to the use of their data—that is, that allows them to engage in research studies and change their consent preferences at any time in narrow and broad ways.[43]

Another option is to have consent models mediated by third parties that negotiate agreements on behalf of data subjects. Data trusts could be modelled after labour unions. This proposal could work quite well, if people were to organize and data experts, ethicists, and lawyers could advise data trusts. Related to consent, if private corporations are to responsibly manage sensitive data about people from which medical information may be inferred, they should have similar fiduciary and confidentiality responsibilities as doctors currently have. If a company betrays the trust of users by misusing their medical data, they should lose their licence to manage medical data.

Companies and institutions should invest in research, software, and infrastructure to manage medical data in the safest way possible. Encryption of data is a must, and security methods such as differential privacy[44]—whereby mathematical noise is inserted into a database to camouflage individual data records—should be implemented. There should also be plans to delete data once it has been used. Only then can patients' consent be valid, as only then will they know how their data will be used. Without an expiry date, data can be used in any way in the future. Data deletion is also an effective way to minimize the risks of misuse.[45]

Apart from making bad practices illegal, giving patients some degree of control over their data through some form of consent, holding corporations to respect fiduciary duties and confidentiality, implementing the best possible security protocols, and deleting data, regulation must ensure that private companies do not monopolize medical data. Above and beyond the risks that can spill outside the doctor's office, medical risks are perhaps the most important to take into account in the context of medical settings. For big data to be a positive force in medicine, we must ensure that all data subjects on whose data it relies can benefit equally from it. Privacy risks are an obstacle to this objective because people can be discriminated against on account of data about them—they can be denied services, or charged more for them. Data risks related to monopolies can also make prices increase, allowing data titans to unduly privately benefit from data they have harvested from the public.

These risks provide a strong argument in favour of having universal health care.[46] The most effective way of protecting people from suffering unfair consequences in medical settings is by having public universal health care coverage in which citizens contribute to the system through taxes and not through dues based on data about them. Such a system would

[42] Mittelstadt and Floridi (n 37).

[43] J Kaye, EA Whitley, D Lund, et al, 'Dynamic Consent: A Patient Interface for Twenty-First Century Research Networks' (2015) 23 European Journal of Human Genetics 141.

[44] C Dwork, 'Differential Privacy' in M Bugliesi et al (eds), *Automata, Languages and Programming* (Springer, 2006), vol 4052.

[45] C Véliz, 'Tus datos son tóxicos', *El País* (8 April 2018), https://elpais.com/tecnologia/2018/04/06/actualidad/1523030681_007734.html accessed 15 July 2019.

[46] There are other arguments in favour of universal health care, but those are outside the scope of this chapter. For an overview of some of these arguments, see N Daniels, 'Justice and Access to Health Care', *Stanford Encyclopedia of Philosophy* (winter 2017 edn), https://plato.stanford.edu/entries/justice-healthcareaccess/ accessed 15 July 2019.

do nothing to prevent work discrimination or extortion resulting from medical data misuses, but it can minimize medical identity theft and, most importantly, it can guarantee that people will not suffer unfairness in access to health care on account of health data monopolies or discrimination.

A strong public and universal health-care system not only provides the most robust protection citizens can get to ensure fair access to medical care—it is also the kind of entity that has enough power to negotiate fair deals with corporations. The NHS, for instance, has enough data about patients, enough medical technology, and enough connections to research institutions to be able to collaborate with corporations without buckling under their weight. In the future, it ought to hire expert lawyers in order to negotiate a better deal with companies like Google so that patients' interests are better protected. Medical data is both very sensitive and very valuable, and health-care institutions should not be giving it out without patients' consent and without ensuring that they will retain some power over the resulting technology.

A. Some Objections

1. The Personal Responsibility Objection

I have argued that privacy risks are dangerous in relation to medical data because they may result in people being treated differently from each other on the basis of personal information about them. But it could be objected that people ought to be treated differently in virtue of how responsible they are for their health conditions. In other words, it could be argued that we should welcome big data analyses even when they come at the cost of losing medical privacy because they will enable accurate attribution of responsibility within health-care settings, so that patients who are less responsible for their disease get more benefits, and patients who are more responsible get less, and that would be fairer than all patients being treated equally even when some of them might be responsible for their own misfortune. However, it will be hard to make sure institutions are not discriminating against people on the basis of information that should not be taken into account, such as genetic make-up, or race. Moreover, the attribution of responsibility is bound to be very controversial. First, algorithms on which big data depend can and do make mistakes.[47] Second, this kind of assessment is not a purely scientific determination, but a value-laden philosophical task that depends on how we conceptualize responsibility.

Furthermore, health is largely a matter of luck: a combination of having the right genes, living in a healthy environment, having had the right education, etc. Cancer, for instance, is the most common cause of death in the world, and a recent study suggests that most cancers are the result of random mutations.[48] Given the weight of luck in health status, an important factor for ensuring all citizens have a fair equality of opportunity is providing universal access to health care. Maintaining health contributes to being able to access the wide range of opportunities available in a given society. If we want to protect equality of opportunity, then

[47] For a good compendium of such mistakes, see C O'Neil, *Weapons of Math Destruction* (Kindle edition, Penguin, 2016).

[48] C Tomasetti, L Li, and B Vogelstein, 'Stem Cell Divisions, Somatic Mutations, Cancer Etiology, and Cancer Prevention' (2017) 355(6331) Science 1330.

offering universal access to health care will contribute to setting off the inequalities that are brought about by sheer luck.

2. The Free Rider Objection

I am proposing two courses of action that may seem in tension with each other. On the one hand, I am proposing asking for informed consent for the collection of medical data, and at the same time, I am defending a universal system of public health care. Believers in desert (ie, roughly, that people should get what they deserve) might want to object that it would be unfair for people who do not give their consent for the collection of their medical data to benefit from a medical system that depends on that data to advance knowledge.

This objection, I suspect, comes from underestimating the cost of donating data; since there is no phenomenological feeling to have one's data collected (it is not a bodily-invasive procedure, it does not physically hurt), it is understandable for someone to think that people who do not donate their data are just being selfish. But the risks of sharing medical data are real, as examples show, and they are likely to become riskier as time goes by and there is more data about people to be aggregated. Donating medical data should thus be thought analogous to participating in clinical research: because it is a risky endeavour, subjects should not be compelled to participate. Participation should always be free, and withdrawal from participation without reprisal must always be an option. Importantly, according to the Declaration of Helsinki, regarded by most as one of the main guidelines for medical research: 'While the primary purpose of medical research is to generate new knowledge, this goal can never take precedence over the rights and interests of individual research subjects.'[49]

People who do not donate their medical data are no worse than people who have not participated in medical research and yet benefit from medical advancements. To recruit data subjects, researchers should do what clinical researchers do to convince people to participate in research: make sure appropriate safety measures are put in place (which, in the case of data, has to include a plan to delete data), ask for informed consent, and give people sufficient compensation to make it worth their while to participate.

3. The Brakes to Innovation Objecton

A third objection to making an effort to protect privacy in medical settings is that limiting the access to and management of medical data will put a brake on data-driven innovation.[50] Ethical limits, however, should not be considered undue obstacles to innovation. In the long run, unethical practices breed distrust and resentment, and a backlash could end up being much more of a barrier to innovation. As an (extreme) analogy, doctors killing healthy people in order to transplant their organs might save lives in the short term, but it will not lead us to the kind of society we want to live in. And it would be a matter of time before people stopped going to the doctor. Privacy is a right for good reasons, and rights are meant to serve as 'red lines' that should not be crossed, even when short-term benefits for violating them might be tempting. Furthermore, as I suggested before, it is still unclear to what extent big data has the capacity to bring about significant medical benefits. The burden

[49] World Medical Association Declaration of Helsinki—Ethical Principles for Medical Research Involving Human Subjects (adopted by the 18th WMA General Assembly 1964) (Note of Clarification added 2002), art 8.

[50] N Price, and G Cohen, 'Privacy in the Age of Medical Big Data', *Nature Medicine* 25, 37–43, 2019.

of proof is on big data to show its worth; in the meantime, especially given the lack of proper regulation, patients should not be forced to gamble with their most sensitive data in exchange for uncertain future benefits.

IV. Conclusion

Big data promises to significantly enhance the power of medicine to diagnose, treat, and prevent diseases. With this promise, however, come significant privacy risks to data subjects who could suffer unfair discrimination, exposure, extortion, and limited access to health care. To minimize these risks, inappropriate uses of data should be outlawed, and consent must be sought from data subjects, even if it is a limited form of consent such as tiered consent. Companies managing sensitive information must also be held to respect fiduciary duties and confidentiality regarding medical data. Security measures like encryption and other cryptographic methods such as differential privacy should be implemented. Data should be deleted after use. Finally, corporations should not be allowed to hold complete power over medical big data. If corporations monopolize medical big data, the best treatments in medicine may only be available to the rich, or to ordinary people under unfair data deals. As things stand, there do not seem to be enough structures to guarantee that public goods and interests will prevail above private interests in the use of big data for medical purposes. If, however, public health-care systems manage to negotiate control over data—through harnessing the weight of the data they have, their connections to research institutions, the trust of their patients (if they manage to keep it), the ability of government to regulate medical settings, and hiring top lawyers to represent the public's interest—fruitful collaborations between the private and public sectors may ensue, and patients may have their privacy and interests better protected. Medical big data can only be successful and ethical if it respects people's right to privacy.

Bibliography

Andreu-Perez, J, Poon, CC, Merrifield, RD, Wong, ST, and Yang, GZ, 'Big Data for Health' (2015) 19(4) IEEE Journal of Biomedical Health and Informatics 1193.

Angus, DC, 'Fusing Randomized Trials with Big Data: The Key to Self-Learning Health Care Systems?' (2015) 314(8) Journal of the American Medical Association 767.

Broniatowski, DA, Paul, MJ, and Dredze, M, 'National and Local Influenza Surveillance through Twitter: An Analysis of the 2012–2013 Influenza Epidemic' (2013) 8(12) PLoS One e83672.

Chunara, R, Bouton, L, Ayers, JW, and Brownstein, JS, 'Assessing the Online Social Environment for Surveillance of Obesity Prevalence' (2013) 8(4) PLoS One e61373.

Daniels, N, 'Justice and Access to Health Care', *Stanford Encyclopedia of Philosophy* (winter 2017 edn), https://plato.stanford.edu/entries/justice-healthcareaccess/ accessed 15 July 2019.

de Montjoye, YA, Hidalgo, CA, Verleysen, M, and Blondel, VD, 'Unique in the Crowd: The Privacy Bounds of Human Mobility' (2013) 3 Scientific Reports 1376.

de Montjoye, YA, Radaelli, L, Singh, VK, and Pentland, AS, 'Identity and Privacy. Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata' (2015) 347(6221) Science 536.

Dwork, C, 'Differential Privacy' in Bugliesi, M, Preneel, B, Sassone, V, and Wegener, I (eds), *Automata, Languages and Programming* (Springer, 2006), vol 4052.

Hern, A, 'Hackers Publish Private Photos from Cosmetic Surgery Clinic', *The Guardian* (31 May 2017), https://www.theguardian.com/technology/2017/may/31/

hackers-publish-private-photos-cosmetic-surgery-clinic-bitcoin-ransom-payments accessed 15 July 2019.

Herring, J, *Medical Law and Ethics* (6th edn, Oxford University Press, 2016).

HIPAA Journal post, 'Largest Healthcare Data Breaches of 2017' (4 January 2018), https://www.hipaajournal.com/largest-healthcare-data-breaches-2017/ accessed 15 July 2019.

Institute of Medicine, Committee on the Role of Institutional Review Boards in Health Services Research Data Privacy Protection, *Protecting Data Privacy in Health Services Research* (National Academies Press, 2000), https://www.nap.edu/read/9952/chapter/1 accessed 15 July 2019.

Kaye, J, Whitley, EA, Lund, D, et al, 'Dynamic Consent: A Patient Interface for Twenty-First Century Research Networks' (2015) 23 European Journal of Human Genetics 141.

Khan, MA-u-d, Uddin, MF, and Gupta, N, 'Seven V's of Big Data: Understanding Big Data to Extract Value' (Proceedings of Zone 1 Conference of the American Society for Engineering Education, 2014), doi: 10.1109/ASEEZone1.2014.6820689.

Kohane, IS and Altman, RB, 'Health-Information Altruists—A Potentially Critical Resource' (2005) 353(19) New England Journal of Medicine 2074–7.

Laney, D, '3D Data Management: Controlling Data Volume, Velocity and Variety' (2001) 6 META Group Research Note.

Mittelstadt, BD and Floridi, L, 'The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts' (2016) 22(2) Science and Engineering Ethics 303–41.

Munro, D, 'Data Breaches in Healthcare Totaled over 112 Million Records in 2015', *Forbes* (31 December 2015), https://www.forbes.com/sites/danmunro/2015/12/31/data-breaches-in-healthcare-total-over-112-million-records-in-2015/ - 5118fabc7b07 accessed 15 July 2019.

Murphy, M, 'Privacy Concerns as Google Absorbs DeepMind's Health Division', *The Telegraph* (13 November 2018), https://www.telegraph.co.uk/technology/2018/11/13/privacy-concerns-google-absorbs-deepminds-health-division/ accessed 15 July 2019.

Nuffield Council on Bioethics, *The Collection, Linking and Use of Data in Biomedical Research and Health Care: Ethical Issues* (Nuffield Council on Bioethics, February 2015).

'Oath of Hippocrates' in WT Reich (ed), *Encyclopedia of Bioethics* (Macmillan, 1995).

O'Neil, C, *Weapons of Math Destruction* (Kindle edn, Penguin, 2016).

Parent, WA, 'Privacy, Morality, and the Law' (1983) 12(4) Philosophy and Public Affairs 269.

Powles, J, 'Why Are We Giving Away Our Most Sensitive Health Data to Google?', *The Guardian* (5 July 2017), https://www.theguardian.com/commentisfree/2017/jul/05/sensitive-health-information-deepmind-google accessed 15 July 2019.

Powles, J and Hodson, H, 'Google DeepMind and Healthcare in an Age of Algorithms' (2017) 7(4) Health and Technology 351, https://link.springer.com/article/10.1007/s12553-017-0179-1 accessed 15 July 2019.

Tanner, A, *Our Bodies, Our Data: How Companies Make Billions Selling Our Medical Records* (Beacon Press, 2017).

Thorpe, JH and Gray, EA, 'Big Data and Public Health: Navigating Privacy Laws to Maximize Potential' (2015) 130(2) Public Health Reports 171.

Tomasetti, C, Li, L, and Vogelstein, B, 'Stem Cell Divisions, Somatic Mutations, Cancer Etiology, and Cancer Prevention' (2017) 355(6331) Science 1330.

Véliz, C, 'On Privacy' (DPhil Thesis, University of Oxford, 2017).

Véliz, C, 'Tus datos son tóxicos', *El País* (8 April 2018), https://elpais.com/tecnologia/2018/04/06/actualidad/1523030681_007734.html accessed 15 July 2019.

Weber, GM, Mandl, KD, and Kohane, IS, 'Finding the Missing Link for Big Biomedical Data' (2014) 311 Journal of the American Medical Association 2479.

World Medical Association Declaration of Helsinki—Ethical Principles for Medical Research Involving Human Subjects (adopted by the 18th WMA General Assembly 1964) (Note of Clarification added 2002).