



NLM Citation: Véliz C. Medical privacy and big data: A further reason in favour of public universal healthcare coverage. In: de Campos TC, Herring J, Phillips AM, editors. *Philosophical Foundations of Medical Law* [Select Chapters]. Oxford (UK): Oxford University Press; 2019.
Bookshelf URL: <https://www.ncbi.nlm.nih.gov/books/>



Medical privacy and big data: A further reason in favour of public universal healthcare 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 good will we depend to avoid abuses of power.² Most 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 minimise those risks. I argue that the most effective way of protecting people from suffering unfair medical consequences is by having a public universal healthcare system in which coverage is not influenced by personal data (e.g., genetic predisposition, exercise habits, eating habits, etc.).

This is a draft of a chapter that has been accepted for publication by [Oxford University Press](#) in the forthcoming book *Philosophical Foundations of Medical Law* edited by Thana C. de Campos, Jonathan Herring, and Andelka M. Phillips due for publication in 2019.

Chapter to be published in: *Philosophical Foundations of Medical Law*, Oxford University Press (2019).

© Carissa Véliz 2019

Monographs, or book chapters, which are outputs of Wellcome Trust funding have been made freely available as part of the [Wellcome Trust's open access policy](#)

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

1

2 Ibid

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 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 (e.g., 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 (e.g., 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 (e.g., intimate sharing of death, illness and recovery); proprietary privacy (e.g., self-ownership and control over personal identifiers, genetic data, and biospecimens); and decisional privacy (e.g., 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 (i.e., about who should profit from patients' genetic data), then it is 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 a) certain kinds of harms that may come about as a result of other people having access to our personal life (e.g., discrimination, identity theft, etc.), b) the demands of sociality, c) being judged and possibly ridiculed by others (and thus from self-conscious negative emotions such as shame and embarrassment), and d) 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

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 (e.g. indisputable/undisputed, inalterable/unaltered, etc.).

5 Carissa Véliz, 'On Privacy' (DPhil thesis, University of Oxford 2017) Chapter 4

6 Anita Allen, 'Privacy and Medicine' in Edward N Zalta (ed), *The Stanford Encyclopedia of Philosophy* (2016) < Available at: <https://plato.stanford.edu/archives/win2016/entries/privacy-medicine/>> last accessed 27 November 2018

7 William A Parent, 'Privacy, Morality, and the Law' (1983) 12 *Philosophy and Public Affairs* 269

8 Véliz (n 5)

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 care, they have little choice but to surrender both sensitive information about themselves and access to their bodies to healthcare 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 healthcare professionals and researchers, however, to minimise privacy losses, to show consideration towards patients and subjects, and avoid any unnecessary exposure.

Thus, healthcare professionals take care to protect patients' autotopos. Sensorial privacy is protected through bedside manners (e.g., looking away while the patient undresses), and other practices designed to limit exposure to patients' autotopos (e.g., 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 stigmatisation and discrimination (e.g., 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 anonymise 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, it should be destroyed).

Anonymisation and informed consent have never been perfect tools for protecting privacy. Genetic data, for example, resists anonymisation because it is unique to each individual. Similarly, in some cases it is impossible to extract valid informed consent (e.g., 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 closed environment (e.g., using computers that are not connected to the Internet, and in a locked room to which only authorised 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 anonymisation and informed consent to the point of risking making them null.

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 Jonathan Herring, *Medical Law and Ethics* (6th edn, Oxford University Press 2016) 224

12 Cited by Anita Allen, 'Privacy and Medicine' (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) < Available at: <https://www.nap.edu/read/9952/chapter/1> > last accessed 27 November 2018

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 revolutionise 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 minimise risks.

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 at 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. Randomised controlled trials (RCTs) have been crucial in the development of rigorous 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 narrow (trial population and setting may not be representative of the general practice); the randomisation 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

13 Doug Laney, '3D data management: Controlling data volume, velocity and variety' (2001) 6 META Group Research Note

14 Laney's conceptualisation has become the classic definition. Subsequent approaches have added other elements to big data. See, for example, M Ali-ud-din Khan, Muhammad Fahim Uddin and Navarun 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/AEEZone1.2014.6820689. 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 emphasise 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 Jane H Thorpe and Elizabeth A Gray, 'Big data and public health: Navigating privacy laws to maximize potential' (2015) 130(2) Public Health Reports 171 PubMed PMID: 25729109.

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 PubMed PMID: 24349542.

17 R Chunara et al, 'Assessing the online social environment for surveillance of obesity prevalence' (2013) 8 PLoS One e61373 PubMed PMID: 23637820.

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 randomisation, and treatment could potentially be much more immediately available.¹⁸

Big data may thus greatly advance personalised 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 decades. 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, '[y]ou 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

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 minimised by technological limits. When health records or data from studies was collected and stored in paper in hospitals and doctors' offices, it was 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 only in the United States.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

18 DC Angus, 'Fusing Randomized Trials With Big Data: The Key to Self-learning Health Care Systems?' (2015) 314 JAMA 767 PubMed PMID: 26305643.

19 J Andreu-Perez et al, 'Big data for health' (2015) 19(4) IEEE J Biomed Health Inform 1193, 1197 PubMed PMID: 26173222.

20 Quoted by Tanner (n 1) 164.

21 Dan Munro, 'Data Breaches In Healthcare Totaled Over 112 Million Records in 2015' *Forbes* (31 December 2015) < Available at: <https://www.forbes.com/sites/danmunro/2015/12/31/data-breaches-in-healthcare-total-over-112-million-records-in-2015/#5118fabc7b07>>

22 HIPAA Journal post, 'Largest Healthcare Data Breaches of 2017' (4 January 2018) < Available at: <https://www.hipaajournal.com/largest-healthcare-data-breaches-2017/>> last accessed 27 November 2018

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 does not only come from traditional sources like EHRs and genetic and microbiomic sequencing data, but also from smartphone applications, data brokers, social networks, Internet searches, environmental data (e.g., 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 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 (e.g., 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 healthcare. 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 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 personalised pricing (i.e., 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 monopolise access to health care. Dominance in data collection leads to

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 JAMA 2479 PubMed PMID: 24854141.

25 Thorpe and Gray (n 15)

26 Julia Powles, 'Why are we giving away our most sensitive health data to Google?' *The Guardian* (5 July 2017) < Available at: <https://www.theguardian.com/commentisfree/2017/jul/05/sensitive-health-information-deepmind-google>>

27 Margi Murphy, 'Privacy concerns as Google absorbs DeepMind's health division' *The Telegraph* (13 November 2018) < Available at: <https://www.telegraph.co.uk/technology/2018/11/13/privacy-concerns-google-absorbs-deepminds-health-division/>>

28 Julia Powles and Hal Hodson, 'Google DeepMind and healthcare in an age of algorithms' (2017) 7(4) *Health and Technology* 351–67 < Available at: <https://link.springer.com/article/10.1007/s12553-017-0179-1>> last accessed 27 November 2018 PubMed PMID: 29308344.

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, because 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 (e.g., 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 (e.g., 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 (e.g., penalise 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 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 personalised medicine with the knowledge and consent of data subjects, protecting people's privacy, and minimising risks that stem from the collection and use of sensitive data. In what follows I explain why anonymisation and consent are necessary but insufficient tools to meet this challenge.

The limits of anonymisation and consent in the context of big data

As was mentioned in Section I, heretofore, medical ethics has relied on anonymisation 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 anonymisation first.

Anonymisation has always been a challenge. Studies that involve photographs, for example, or genetic data, can be hard or impossible to anonymise.³² But, for the most part, other kinds of data could be more easily anonymised by stripping away information that could identify individuals. The use of big data, however, makes anonymisation 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 electronic 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

29 Ibid

30 Alex Hern, 'Hackers publish private photos from cosmetic surgery clinic' *The Guardian* (31 May 2017) < Available at: <https://www.theguardian.com/technology/2017/may/31/hackers-publish-private-photos-cosmetic-surgery-clinic-bitcoin-ransom-payments>>

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, Mandl and Kohane (n 24)

Therefore, every person must be uniquely tagged with a medical identifier, making it all the more easier to re-identify individuals.

Even if it were possible to anonymise data in a secure way, the analysis of aggregated data can also result in group-level harms such as stigmatisation 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 anonymisation 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 kind of correlations may be unveiled, and they often cannot guarantee how this data will be used.³⁸

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

Anonymisation 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 minimised? I turn to this question next.

III. Minimising risks

Part of what made anonymisation 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 minimise 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 anonymised 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 (for example, by inferring sensitive information from non-sensitive information, or by storing data in a way that could allow for

34 YA de Montjoye et al, 'Unique in the Crowd: The privacy bounds of human mobility' (2013) 3 Sci Rep 1376 PubMed PMID: 23524645.; YA de Montjoye et al, 'Identity and privacy. Unique in the shopping mall: on the reidentifiability of credit card metadata' (2015) 347 Science 536 PubMed PMID: 25635097.

35 Tanner (n 1) 91

36 Andreu-Perez et al (n 19) 1204

37 Brent Daniel Mittelstadt and Luciano Floridi, 'The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts' (2016) 22(2) Science and Engineering Ethics 303 PubMed PMID: 26002496.

38 Ibid

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 27 November 2018

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 consent for each use of their data. This approach is incredibly burdensome, however, both for institutions and individuals. It is too impractical to ask for consent to every individual for every single use of data from every source. 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 recall their data. A more promising proposal is 'tiered' consent, in which data subjects can choose specific future uses of their data (e.g., 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 organise and data experts could advise data trusts. In terms of feasibility, however, it is unclear whether there is enough awareness of privacy issues in our societies for people to be willing to organise and defend their rights. But perhaps as privacy-related incidents such as identity theft and extortion continue to rise, data subjects will take steps to organise themselves into groups that can better stand up for their rights against institutions.

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 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 privacy⁴⁴—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 minimise the risks of misuse.⁴⁵

41 IS Kohane and RB Altman, 'Health-information altruists--a potentially critical resource' (2005) 353(19) N Engl J Med 2074 PubMed PMID: 16282184.

42 Mittelstadt and Floridi (n 37)

43 Jane Kaye et al, 'Dynamic consent: a patient interface for twenty-first century research networks' (2015) 23 European Journal of Human Genetics 141 PubMed PMID: 24801761.

44 Cynthia Dwork, 'Differential Privacy' in Michele Bugliesi et al (eds), *International Colloquium on Automata, Languages, and Programming Automata, Languages and Programming*, vol 4052 (Springer 2006)

45 Carissa Véliz, 'Tus datos son tóxicos' *El País* (8 April 2018) < Available at: https://elpais.com/tecnologia/2018/04/06/actualidad/1523030681_007734.html>

Apart from making bad practices illegal, giving patients some degree of control over their data through some form of consent, holding corporations to respect confidentiality, implementing the best possible security protocols, and deleting data, regulation must ensure that private companies do not monopolise medical data. Above and beyond the risks that can spill outside the doctors' 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 to improve 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 do nothing to prevent work discrimination or extortion resulting from medical data misuses, but it can minimise medical identity theft, and most importantly, it can guarantee that people will not suffer unfairness in access to healthcare 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 health 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 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.

Some objections

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 healthcare 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 people on the basis of information that should not be taken into account, such as genetic makeup, 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 conceptualise 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

46 There are other arguments in favour of universal healthcare, but those are outside the scope of this paper. For an overview of some of these arguments, see Norman Daniels, 'Justice and Access to Health Care.' *Stanford Encyclopedia of Philosophy*.

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

world, and a recent study suggests 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 healthcare. 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 offering universal access to healthcare will contribute to setting-off the inequalities that are brought about by sheer luck.

The free rider objection

I am proposing two things 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 healthcare. Believers in desert 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 nothing it feels like to have one's data collected, 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, '[w]hile 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 enough compensation to make it worth their while to participate.

IV. Conclusion

Big data promises to significantly enhance the power of medicine to diagnose, treat, and prevent of diseases. With this promise, however, come significant privacy risks to data subjects who could suffer unfair discrimination, exposure, extortion, and limited access to healthcare. To minimise 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 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 monopolise 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 healthcare systems manage to negotiate power over big 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), and the ability of government to regulate

48 C Tomasetti, L Li, and B Vogelstein, 'Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention' (2017) 355(6331) *Science* 1330 PubMed PMID: 28336671.

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) article 8

medical settings, fruitful collaborations between the private and public sectors may ensue, and patients may have their privacy and interests better protected.

Acknowledgements

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

Bibliography

- Andreu-Perez J and others, 'Big data for health' (2015) 19(4) IEEE J Biomed Health Inform 1193–208. PubMed PMID: 26173222.
- Angus DC, 'Fusing Randomized Trials With Big Data: The Key to Self-learning Health Care Systems?' (2015) 314(8) JAMA 767–8 PubMed PMID: 26305643.
- 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 PubMed PMID: 24349542.
- Chunara R and others, 'Assessing the online social environment for surveillance of obesity prevalence' (2013) 8(4) PLoS One e61373 PubMed PMID: 23637820.
- Daniels N, 'Justice and Access to Health Care.' Stanford Encyclopedia of Philosophy.
- de Montjoye YA and others, 'Unique in the Crowd: The privacy bounds of human mobility' (2013) 3 Sci Rep 1376 PubMed PMID: 23524645.
- de Montjoye YA et al, 'Identity and privacy. Unique in the shopping mall: on the reidentifiability of credit card metadata' (2015) 347(6221) Science 536 PubMed PMID: 25635097.
- Dwork C, 'Differential Privacy' in Bugliesi M et al (eds), *International Colloquium on Automata, Languages, and Programming Automata, Languages and Programming*, vol 4052 (Springer 2006)
- Hern A, 'Hackers publish private photos from cosmetic surgery clinic' *The Guardian* (31 May 2017) < Available at: <https://www.theguardian.com/technology/2017/may/31/hackers-publish-private-photos-cosmetic-surgery-clinic-bitcoin-ransom-payments>> last accessed 27 November 2018
- HIPAA Journal post, 'Largest Healthcare Data Breaches of 2017' (4 January 2018) < Available at: <https://www.hipaajournal.com/largest-healthcare-data-breaches-2017/>> last accessed 27 November 2018
- Herring J, *Medical Law and Ethics* (6th edn, Oxford University Press 2016)
- 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) < Available at: <https://www.nap.edu/read/9952/chapter/1>> last accessed 27 November 2018
- Kaye J et al, 'Dynamic consent: a patient interface for twenty-first century research networks' (2015) 23 European Journal of Human Genetics 141–6. PubMed PMID: 24801761.
- 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) N Engl J Med 2074–7 PubMed PMID: 16282184.
- 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 PubMed PMID: 26002496.

- Munro D, 'Data Breaches In Healthcare Totaled Over 112 Million Records in 2015' *Forbes* (31 December 2015) < Available at: <https://www.forbes.com/sites/danmunro/2015/12/31/data-breaches-in-healthcare-total-over-112-million-records-in-2015/-5118fab7b07>> last accessed 27 November 2018
- Murphy M, 'Privacy concerns as Google absorbs DeepMind's health division' *The Telegraph* (13 November 2018) < Available at: <https://www.telegraph.co.uk/technology/2018/11/13/privacy-concerns-google-absorbs-deepminds-health-division/>> last accessed 27 November 2018
- 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 edition, 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) < Available at: <https://www.theguardian.com/commentisfree/2017/jul/05/sensitive-health-information-deepmind-google>>
- Powles J and Hodson H, 'Google DeepMind and healthcare in an age of algorithms' (2017) 7(4) *Health and Technology* 351–67 < Available at: <https://link.springer.com/article/10.1007/s12553-017-0179-1>> last accessed 27 November 2018 PubMed PMID: 29308344.
- 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 PubMed PMID: 28336671.
- Véliz C, 'On Privacy' (DPhil Thesis, University of Oxford 2017)
- Véliz C, 'Tus datos son tóxicos' *El País* (8 April 2018) < Available at: https://elpais.com/tecnologia/2018/04/06/actualidad/1523030681_007734.html>
- Weber GM, Mandl KD and Kohane IS, 'Finding the missing link for big biomedical data' (2014) 311 *JAMA* 2479 PubMed PMID: 24854141.
- 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)