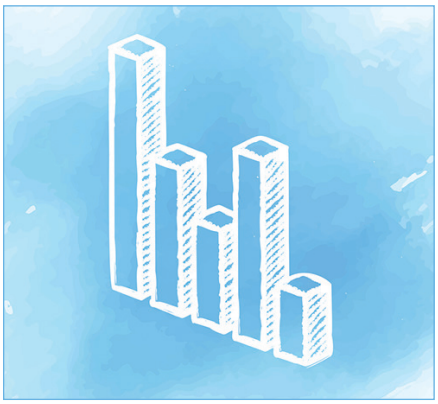

Ethical Challenges Posed by Big Data

by EDMUND G. HOWE, III, MD, JD, and FALICIA ELENBERG, BS

Dr. Howe is a Professor of Psychiatry and Medicine, Director of Medical School Programs in Ethics, and Senior Scientist at Uniformed Services University of the Health Sciences in Bethesda, Maryland.

Innov Clin Neurosci. 2020;17(10–12):24–30

ABSTRACT

Big Data is a term that refers to tremendously large data sets intended for computational analysis that can be used to advance research through revealing trends and associations. Innovative research that leverages Big Data can dramatically advance the fields of medicine and public health but can also raise new ethical challenges. This paper explores these challenges, and how they might be addressed such that individuals are optimally protected. Key ethical concerns raised by Big Data research include respecting patient's autonomy via provision of adequate consent, ensuring equity, and respecting participants' privacy. Examples of actions that could be taken to address these key concerns on a broader regulatory level, as well as on a case specific level, are presented. Big Data research offers enormous potential, but due to its widespread influence, it also introduces the potential for extensive harm. It is imperative to consider and account for the risks associated with this research.

KEYWORDS: Big Data, data privacy, healthcare, artificial intelligence, ethics

The emergence of Big Data as a tool can exponentially advance our medical care.[1] Big Data is defined as consisting of tremendously large data sets intended for computational analysis that can be used to advance research through revealing new trends and associations.[2] A recent and somber example of how Big Data can revolutionize health is by providing us with early knowledge regarding emerging pandemics.[3] More broadly, Big Data research can provide insights that help us better understand social determinants of health, discover novel treatments, and map the underlying mechanisms, markers, and progression of diseases. An illustration of these capabilities are advancements made in the field of medical imaging as a result of the application of artificial intelligence (AI) to Big Data. Algorithms trained using massive medical image datasets have led to dramatically enhanced abilities to detect diabetic retinopathy and skin cancer.[4,5] Diagnostic accuracy has been improved to the extent that computers can now surpass even clinicians' capacity to disentangle complex and subtle discriminative patterns.[6] There are countless more examples of current benefits derived from Big Data, and its use in medical research will only increase.

Despite the many gains, Big Data research has also raised new ethical concerns. These concerns partially stem from the fact that Big Data research requires access to large scale amounts of personal information. They also stem from the tendency of analytics programs to reflect human error. In January of 2019, the Revised Common Rule, a set of regulations governing federally supported human subject research, was adopted. While these new standards provided many needed updates to research requirements in the wake of the digital age, they inadequately attended to many aspects of the specific concerns posed by Big Data research.[7]

The Revised Common Rule inadequately attends to the challenges raised by Big Data research because it treats Big Data research in a similar manner to traditional research. The concern about this approach is that Big Data research is drastically different—it leverages new sources of information and methodologies and produces novel types of results. Unlike traditional research, Big Data research often relies on very large sets of publicly available information. However, many of the regulations in the Revised Common Rule, such as clear informed consent, do not apply to this type of data. In general, it is believed that there is less of a need to protect publicly available information. This has resulted in participants being left unaware of the use, or purpose of use, of their information. Lack of stronger regulations regarding publicly available data has also left people more vulnerable to re-identification and other privacy threats.[8] Further concerns that have stemmed from current uses of Big Data include issues centering around bias and equity. Big Data is changing how studies are conducted. To sufficiently prepare for the shift toward utilizing this research approach, these changes have to be carefully considered.

We can, and should, better prepare for the consequences of Big Data research. An

FUNDING: No funding was provided for this study.

DISCLOSURES: The opinions or assertions contained herein are the private views of the authors and are not necessarily those of the AFRRI, USUHS, or the Department of Defense.

CORRESPONDENCE: Edmund G. Howe, III, MD, JD; Email: edmund.howe@usuhs.edu

analogous, historical example involves the Human Genome Project. In anticipation of the results from this study, the Ethical, Legal, and Social Implications (ELSI) program was founded to identify and address ethical issues that could arise. They were tasked with positing how the project could affect individuals at both the societal and personal levels, and then creating recommendations that would facilitate the positive effects while minimizing the negative ones.[9,10] Similarly, we could anticipate ethical concerns that uses of Big Data might give rise to and enact programs meant to protect participants.

Uses of Big Data can have a significant effect on people's health, both positive and negative. This piece will provide an overview of some of the main ethical challenges uses of Big Data could pose, with a specific focus on the field of medicine. The discussion below will be organized according to three primary ethical principles: respecting patients' autonomy, maintaining equity, and protecting privacy. Potential approaches for addressing these ethical challenges will then be discussed. The goal of this endeavor, as stated flawlessly by Dr. Mittelstadt and Professor Floridi of the Oxford Internet Institute, is as follows: "Ethical foresight may reduce the probability of 'regulatory whiplash' by informing public debate through improved understanding of the 'moral potential' of emerging technological applications and data practices."[11]

ETHICAL PRINCIPLES CHALLENGED BY BIG DATA RESEARCH

The three ethical concepts that might most likely be challenged by uses of Big Data are those previously mentioned—respecting participants' autonomy, achieving equity, and protecting privacy. These goals are shared by all Institutional Review Boards (IRBs) seeking to protect participants from the foreseeable harms that research could bring about. Each risk must be individually considered and then must be weighed against Big Data's anticipated contribution to people's health and wellbeing.

Respecting patients' autonomy. The principal approach taken in research to respect individuals' autonomy is obtaining informed consent. The main goal of the consent process in research is to ensure that patients receive an understanding of the purpose, risks, and methodology of the research being conducted.[12] The consent process upholds ethical principles of autonomy and freedom of choice by allowing patients to make a well-informed decision. While the consent process is arguably imperfect and limited, it has been the best available tool for maintaining the aforementioned ethical principles.[13] However, given that Big Data research differs drastically from traditional research in terms of subject participation, one of the biggest concerns is whether current research requirements regarding consent still adequately protect patients' autonomy.

The Revised Common Rule emphasizes the importance of proper consent but does not adequately extend its reach to Big Data research. The new standards emphasize the importance of proper consent by requiring researchers to succinctly state to patients what they need to know in regard to why they would or would not want to participate in a protocol are the very beginning of every informed consent document.[7] Before the enactment of this rule, the voluminous and dense nature of informed consent documents made it difficult for patients to gain an understanding of the risks they would face by participating in a given study. Despite these efforts to improve and uphold consent in traditional research, the Revised Common Rule leaves an avenue for avoiding informed consent in Big Data research; the Revised Common Rule requires only broad consent when publicly available information is used, and no consent is required when de-identified information is used.[14] Broad consent and lack of consent means that participants are not being provided a complete understanding of the uses of their data.

The issue with broad consent is that unlike informed consent documents, which are tailored for a specific project or research use, broad consent documents are more general and relate to an unspecified range of future research studies. It is difficult to make an informed decision about participating in a study when there is strong ambiguity regarding how information will be used. In addition to concerns regarding broad consent, there are also concerns regarding no consent. When Big Data researchers are using de-identified publicly available information, no consent is required from the research participants. Considering that information used in Big Data research is often created by individuals for purposes other than research, individuals might likely be unaware of this potential use of their information. Current consent models are ill suited for the conditions under which information used in Big Data research is created.[15]

Lack of stronger consent requirements for publicly available information generates concerns regarding autonomy, because individuals that generate this data do not often realize the extent to which their information is public. Moreover, individuals might not realize the extent to which their information can be used by others without their permission or the type of inferences can be drawn from analyzing their data. This can easily occur when researchers derive their data sets directly from the internet through platforms such as social media, for example.[11] The websites that Big Data researchers use as sources of information often serve as integral parts of people's daily lives, and a significant proportion of people might regard their information as private despite its accessibility. For example, when someone posts to Facebook about an illness, even though they chose to make their profile viewable by others, they might expect that only those connected to their social circle will see their sensitive post. Not only might they only expect their social circle to see the post, they might also expect that no one will use the post. Current research regulations define public versus private information based on accessibility. This is not necessarily congruent with how the participants define public versus private information.

Even if participants are aware that their information is considered publicly available and can be used in research, they might be unaware of the type of research and findings that can result from use of their information. One of the biggest challenges in respecting participants' autonomy as it pertains to research involving Big Data is the fact that inherent to Big Data is finding unexpected correlations, associations, and trends. As Michael Froomkin, a Professor of Law states, ". . . it follows that neither the researcher nor the subject might know what the data collected will be used to discover."[14] Under this system, even if people appreciate that their information is public and can be used for research, they cannot appreciate in advance what these specific uses might be.

Since it is difficult to know what the information collected for Big Data research

will be used to discover, people might easily and unwittingly participate in research to which they are morally opposed. For example, there was a study conducted to demonstrate the dangers of Big Data that used publicly available photographs. The researchers in this study found pictures of people posted on dating websites and recorded the images of over 70,000 individuals. They then used these photos to train an analytics model to predict sexual orientation based on facial features. It was found that that the AI technology trained using this dataset can in fact use facial features to distinguish between people who identify as gay and heterosexual much more accurately than people.[16] This analytics capability can be used for nefarious purposes against the lesbian, gay, bisexual, and transgender (LGBT) community. Thus, it is fair to assume that not only do many people not realize that their pictures could give away this type of insight, but many people would not want the pictures they post to be used to inform such an analytical model.

It should be noted that the Revised Common Rule does outlaw federally supported entities from carrying out harmful research, but a study does not have to be explicitly harmful for people to not want to participate in it for any number of reasons including moral oppositions. The study above is just an example that proves how easily people's autonomy can be disregarded when research leverages publicly available information. The researchers of the sexual orientation study assert that the fact that Big Data analytics can result in such types of inferences to be made should warrant the utmost concern.

As highlighted in the previous paragraph, many participants might be unaware of the capabilities of analytics. Many people are likely unaware of the fact that facial features have a degree of correlation to sexual orientation, and that AI technologies can recognize this. As technological capabilities rapidly advance, it becomes more and more likely that people will be unable to foresee the ways in which their publicly available information could be used. In almost all medical research, informing participants so they can choose whether or not to be in a protocol is an absolute requirement, since anything less than this is viewed as an unacceptable violation of human dignity.[7] With current regulations allowing for lenient consent requirements for Big Data research, these violations of human dignity seem eminent.

A balanced remedy that fosters innovative Big Data research and maintains the autonomy and dignity of participants is urgently needed. A potential remedy would need to address the concern that the vast majority of Americans feel they have little to no control over data collected about them. It would also need to address the concern that the majority of Americans feel they have very little to no understanding about what government institutions do with data collected about them.[17]

A potential remedy could require prominent warnings on websites, apps, and social media platforms about what is considered publicly available information and how it can be used, such that people can make informed decisions before generating content. The warnings would have to be extremely concise and in an important location, because, as previous studies have shown, participants tend not to read fine print.[18] Schools could potentially take initiative by informing students about how their information can be used in Big Data research. Healthcare providers and government institutions could engage in similar initiatives aimed at better informing the public. Lastly, requirements could be put in place to automatically notify an individual when their information was used in research, with the name of the study included. Many people might ignore the notifications, or opt out of them entirely, but at least that would be participants' decision to make.

Achieving equity. Big Data research poses a challenge to the ethical principal of achieving equity because its results can easily and inadvertently perpetuate disparities. Big Data research can help overrepresented populations, while not providing gains to, and even possibly harming, underrepresented populations. This occurs when Big Data research uses data predominantly obtained from a single group —based on race, ethnicity, country of origin, or socioeconomic class. The conclusions these studies arrive at reflects these participants' characteristics and therefore tend to primarily benefit this one group. When the study's findings are applied to other groups, the benefits might not translate.[19] Moreover, people can be harmed when irrelevant findings are applied to underrepresented populations. For example, a particular treatment that works for one group of people might cause adverse side effects in another. In addition to the problems posed by unrepresentative datasets, another way that Big Data research can challenge equity is through algorithms trained using representative, but biased data. Algorithms trained using biased data produce biased conclusions. Healthcare disparity is among the most significant problems facing the field of medicine today. When used properly, Big Data research might mitigate this problem and promote equity, but when used improperly, Big Data research can perpetuate the already harmful disparities that exist.

A current example of homogenous datasets threatening equity involves the field of genetics. Genetic data is typically produced by either individuals with high quality health insurance plans or consumers with disposable income. As a result, the genetic data that currently exists disproportionately represents individuals with higher income. In fact, certain genetics studies have been found to use data from primarily "Euro-Americans of middle to upper socioeconomic status".[20] Findings from these studies are likely less applicable to individuals who fall outside of this group. The concern that then arises is that use of this homogenous data might exacerbate already existing gaps in medical knowledge and practices.

Even if a data set used in Big Data research is representative of all populations, the results might still be inequitable if input data is biased. An example of biased data sets leading to biased results involves Big Data research used to assist judges in sentencing offenders. The research was meant to assist in judges' sentencing by using an algorithm to predict offenders' likelihood of recidivism. The data sets used to train the algorithm were representative but were tainted by racial discrimination. In America, Black individuals are more likely to be sentenced to jail for a given crime than their white counterparts who commit the same crime. Since the algorithm is trained using data created within this discriminatory context, the algorithm will inherently output that Black individuals have a higher risk of recidivism. In short, the algorithm's conclusions were found to perpetuate the bias it was fed.[21] While the medical ramifications of this study are more remote than in the genetics example, they are still significant. For example, incarceration

often causes acute and chronic stress, both for the individuals being incarcerated as well as their dependents. These types of stress are associated with negative health outcomes, such as immune dysfunction.[22] This example highlights the extent to which bias in Big Data research could cause harm.

A similar, more specific example of how bias in Big Data research can/is perpetuating inequities in healthcare involves an algorithm used to refer patients to care management programs. In October 2019, researchers from the University of California Berkeley, the University of Chicago Booth School of Business, and Partners HealthCare found that an algorithm used to refer patients who were high-risk was perpetuating racial biases. The analytics platform was referring healthier white patients to care management programs at higher rates than it was referring less healthy black patients to those same care management programs. In many areas of the United States providers deliver more care to white patients than their Black counterparts. The result is higher average healthcare dollars spent on white patients with comparable medical conditions. Because the algorithm discussed earlier assesses risk based on the amount of healthcare dollars spent on a given patient, it will tend to refer white patients more often. The researchers detected these inequities by comparing the algorithm's current risk analysis with an analysis of other markers of health risk, such as number of chronic illnesses treated in a year and avoidable cost.

When adjusting the algorithm for those new markers of health risk, the number of black patients referred to care management programs increased from 18 percent to 47 percent of all patients.[23,24] Whether algorithms perpetuate or mitigate inequity is entirely up to us. Regulation and oversight over the choices made when an algorithm is trained can make a big difference. This particular example is so profound because it illustrates the extent to which Big Data research can cause widespread harm; the category of algorithms discussed influence healthcare decisions made for millions of Americans.

Policies that minimize the risks Big Data research poses to equity must be considered. For example, when algorithmic analyses are designed, safeguards can be built in to compensate for known biases. Additionally, areas where increased research is necessary to overcome gaps in knowledge and practices for underrepresented populations should be identified. Addressing the concern of representative data sets might be more difficult, since Big Data research is often unable to "account for those who participate in the social world in ways that do not register as digital signals."[25] In this case, researchers might have to invest in finding and explicitly collecting data from groups who tend to produce digital signals less often. While this investment might be costly to the entity conducting research, it could play a large role in mitigating future healthcare disparities.

An additional concern regarding equity that might arise from uses of Big Data is compensation. Entities doing research on certain populations might fail to share gains of their research with the population they use in their study, and those entities might also not compensate their participants adequately. Previous examples of pharmaceutical research conducted on populations in developing counties illustrates this concern. A potential remedy to this problem might be formal benefit-sharing agreements made between data providers and researchers to help ensure greater equity for research participants.[10]

Protecting privacy. As discussed throughout this paper, Big Data research can potentially pose a wide range of risks to individuals.[26] The harmful risks we shall consider in this section are violations to people's privacy. Privacy is an important and fundamental right of United States citizens. This right is alluded to in the fourth amendment of the Constitution, and it was established as a Constitutional doctrine following the Supreme Court case of Griswold verses Connecticut.[27] The reason why privacy is awarded this level of importance is because it serves to protect people's fundamental liberty. Privacy accomplishes this by serving as a limit to government and corporate power, affording individuals greater control over their lives and the decisions made about them, and protecting individuals from exploitation.[28–30]

There are countless examples of ways in which Big Data research can threaten privacy. Big Data research can "quickly take on surveillance implications,"[31] implications that are inherently incongruent with privacy. One group of experts in this area even go so far as to say that "Big Data has been compared with an omniscient 'transparent human' capable of mass surveillance."[32] This surveillance capability was demonstrated in the sexual orientation study discussed earlier. It is also illustrated in cases where algorithms are used to aid in employment processes. In one study, for example, researchers used AI to scan thousands of babysitters' profiles available on Facebook, Twitter, and Instagram. The researchers then used data analytics to rate the "risk" to children these babysitters posed. Parents were able to access these ratings and use them to decide whether or not to hire the babysitters.[33] AI can similarly analyze people's speech and even their facial expressions to come up with negative appraisals of their trustworthiness and competence.[34] These babysitters never gave consent to their data being analyzed in such a matter, nor were they informed that this was occurring. Moreover, it is possible that many of these babysitters did not even know that researchers possessed the tools to automatically scan through their profiles and generate this type of risk score. While this type of research is not federally supported and falls beyond the scope of the Revised Common Rule, one can imagine that this type of analysis might one day be used to predict how risky it is for a medical provider to take on a certain patient. It is important to anticipate such concerns and plan for them accordingly.

Another way that Big Data research can threaten privacy is through the use of de-identified data. This risk is already beginning to manifest, as can be seen in research involving genetic samples and genetic data. Research participants are often assured that their genetic data will not be identifiable. However, "de-identified" genetic data is by and large a false notion. Researchers have shown that for the vast majority of Americans, de-identified genetic data can be reattached to the identity of the person who provided the initial samples. This reattachment of identity to data is accomplished via family maps created by public genealogy databases.[35] In fact, one curious researcher set out to see how hard it would be to find the identity of a participant in a study using their de-identified genetic data. This researcher had essentially no experience tracking people and was therefore far from an expert at this task. Nonetheless, the researcher found that it was possible to identify multiple

study participants whose data had been claimed to be anonymized. The researcher was able to accomplish this task in just one day by matching the de-identified genetic data with genealogy maps from an open-source site, GEDmatch, as well as by using other tools, such as social media.[36] The vast amount of publicly available information about a given person means that more and more "de-identified" data will become re-identifiable. De-identified data is afforded the fewest protections under the Revised Common Rule, because if information cannot be tied to an individual, it is much less likely to cause them harm. However, given that identity can be reattached in certain circumstances, enhanced guidelines for de-identified data are needed.

MOVES TOWARD POSSIBLE SOLUTIONS

Several examples throughout this manuscript have illustrated the need to enact solutions for the risks posed by Big Data research. These solutions are needed for research that both falls within the scope of the Revised Common Rule and research that does not. Independent efforts to reduce risks stemming from Big Data research have already begun. The Federal Trade Commission (FTC) has investigated private companies based on potential misuse of their consumers' data.[37] Additionally, several consumer organizations have gone further, seeking through Congress to create a new federal agency that is devoted to establishing new privacy protections.

Some organizations have also taken initiative and begun making changes of their own. As an example, Facebook's chief executive has publicized new steps they are taking to increase the privacy of their users. These steps include the use of encryption techniques to protect what users view and send. Corporate efforts such as these to protect users' privacy might be motivated by financial interests, but they can still serve as an example for how research institutions might be able to better handle and protect their subjects' data.[38,39]

An additional solution for mitigating the ethical risks posed by Big Data involves determining the people most at risk in advance of initiating research. This solution would be difficult and serve as a preeminent challenge, but it is feasible. Focus groups are already underway trying to help ascertain who these at-risk groups might be so that their interests can be protected and prioritized.[40]

In addition to protecting privacy and at risks groups, regulations could also be enacted for the purposes of maintaining equity. Regulations can be put in place requiring Big Data research to use representative study samples consisting of individuals from different races, socioeconomic statuses, religions, and ethnicities. Similar mandates in the past have resulted in better representation of both sexes in healthcare research.[41] Justification of any discrimination involving gender in research is now required, and as a result, women have been increasingly included in studies as research subjects. Similar large-scale efforts to achieve proper representation for other groups in healthcare research is needed. An example of such an effort already underway is the Observational Health Data Sciences and Informatics project. This project is an international open science program that includes 1.26 billion patient records from 17 countries and uses a common data model.[19] This platform allows researchers to collect samples from people with varying characteristics and can serve as a tool for achieving greater equity in Big Data research. The next important step is to ensure that these tools are broadened, promote participant privacy, and become widely adopted in research.

Not only can participants benefit from the updating of regulatory standards, but they can also benefit from their scope being expanded. There are many private companies who conduct healthcare research without federal support, and therefore, without strong governance. This results in participants facing greater risk. As an analogous example, the Health Insurance Portability and Accountability Act (HIPAA) only applies to covered entities—health plans, healthcare clearing houses, and healthcare providers that transmit information in connection with a transaction for which United States Department of Health and Human Services (HHS) has adopted a standard.[42] This means that if an individual gets a genetic test done through their doctor's office, their genetic information is protected by HIPAA. But if an individual gets the same genetic test done through a direct-to-consumer company, their genetic information is not protected by HIPAA and is much more vulnerable.

A similar phenomenon exists for research that does not fall under the purview of the Revised Common Rule. As of 2018, roughly 8.1 billion dollars in venture capital was allocated to healthcare digital startups. This was based on the premise that these developing models could bring about extraordinary benefits.[43] These digital startups are going to have significant impacts on healthcare, but their research is not subject to the same level of regulation as compared with federally support human subject research. The scope of the Revised Common Rule could either be extended, or separate, stringent standards for nonfederally funded research could be enacted. Stronger government oversight for private corporations would enable more adequate protection for research participants.[2] Furthering the three ethical principles discussed in this manuscript should be foremost among such standards' chief concerns.

Other remedies at the level of policymaking might emerge, but optimal practices and their implementation will take time. In the meantime, health providers might want to consider how to best inform their individual patients of potential Big Data risks, if they believe they should inform them at all. The risks patients face due to information that they create and post on the internet might warrant at least a brief discussion. The studies mentioned above involving sexual orientation and employment are paradigmatic examples. Even the briefest discussions between healthcare providers and their patients, however, will take time. Would this be worth it? One approach to answering this question is to ask patients what they would prefer—are they interested in learning more about the risks they face as it pertains to Big Data research? Once this question is asked, the decision of whether providers should spend time discussing Big Data research risks would be shared.[44]

At this time, other risks from Big Data research and its applications might still be largely unknown. AI and machine learning techniques are beginning to outstrip our capacity to understand how conclusions are reached.[45] Digital computer programs can arrive at conclusions using incredible speed and complexity, such that even their developers cannot understand how the inputs were used by the program to arrive at a given conclusion.[46] These algorithms have

therefore raised the question of what new risks might be posed when human intelligence can no longer comprehend the pathways of computer-based decisions, otherwise known as the "black box" problem.[47] This will become a particularly prominent concern once machines can teach themselves novel skills.[48] Big Data research that contains within it "black boxes" would be highly problematic since this would preclude the study's methodologies from needed transparency and oversight. Both researchers and the public must have sufficient understanding of how findings have been reached to be able to make moral judgments regarding how new information should be used, if it should be used at all. "Black boxes" already exist in private industry, and fortunately, technology intended to lift these black boxes is currently being created. The black box risk is potentially so significant in healthcare that it warrants ongoing scrutiny, and it warrants federal support for countermeasure development.

CONCLUSION

Optimal ethical solutions should be sought on both a societal and inter-personal level. Governments should especially seek to ensure that persons vulnerable to becoming unwitting, or even witting research participants understand the risks they face.[49] This is especially important because Big Data studies might affect stigmatization, negatively target individuals, and even affect people's livelihoods.

Big Data analytic technologies are tools. Tools are not inherently good or bad but uses of tools can create harmful or beneficial effects. This manuscript is by no means arguing that Big Data research is bad and should be avoided. Big Data research can provide, and already has provided, tremendous benefits to the medical field and to society as a whole by offering invaluable insights. However, the harmful effects that can arise from this tool carry a heavy weight. As such, these risks should be anticipated, and measures should be put in place to reduce their likelihood and impact. As a society and as individual care providers, we must seek to empower patients in face of these extraordinarily fast-paced technological developments. We need to provide our patients with the proper knowledge and tools to make informed decisions, and we need to try to deter harmful ethical outcomes.

REFERENCES

1. Katal A, Wazid M, Goudar RH. Big data: Issues, challenges, tools and good practices. Presented at the Sixth International Conference on Contemporary Computing (IC3); 2013; Noida, Indiana.
2. Emanuel EJ, Wachter RM. Artificial intelligence in health care. *JAMA.* 2019;321(23):2281–2282.
3. Milinovich GJ, Magalhães RJS, Hu W. Role of Big Data in the early detection of ebola and other emerging infectious diseases. *The Lancet Glob Health*. 2015;3(1):e20–21
4. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* 2016;316(22):2402–2410.
5. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;546(7660):1317–1318.
6. Beam AL, Kohane IS. Big Data and machine learning in health care. *JAMA.* 2018;319(13):1317–1318.
7. Menikoff J, Kaneshiro J, Pritchard I. The common rule, updated. *N Engl J Med.* 2017;376(7):613–615.
8. Revised Common Rule, § ___.116(a)(5), 82 Fed. Reg. 7265/3 (codified at 42 C.F.R. § 46.116(a)(5))
9. McGuire AL, Caulfield T, Cho MK. Research ethics and the challenge of whole-genome sequencing. *Nat Rev Genet.* 2008;9(2): 152–156.
10. Mathaiyan J, Chandrasekaran A, Davis S. Ethics of genomic research. *Perspect Clin Res.* 2013;4(1):100.
11. Mittelstadt BD, Floridi L. The ethics of Big Data: current and foreseeable issues in biomedical contexts. *Sci Eng Ethics.* 2016;22(2):303–341.
12. Paterick TJ, Carson GV, Allen MC, Paterick TE. Medical informed consent: General considerations for physicians. *Mayo Clin Proc.* 2008;83(3):313–319.
13. Abujarad F, Alfano S, Bright TJ, et al. Building an informed consent tool starting with the patient: The patient-centered virtual multimedia interactive informed consent (VIC). *AMIA Annu Symp Proc.* 2017;2017:374–383.
14. Froomkin MA. Big Data: Destroyer of informed consent. *Yale J Health Pol'y L & Ethics Spec*. 9 2019 Sept;122.
15. Vayena E, Mastroianni A, Kahn J. Caught in the web: Informed consent for online health research. *Sci Transl Med.* 2013;5(173):173fs6.
16. Wang Y, Kosinski M. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *J Pers Soc Psychol.* 2018;114(2):246–257.
17. Auzier B, Rainie L, Anderson M, et al. Americans and privacy: concerned, confused, and felling lack of control over their person information. 2019. Pew Research Center. https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/. Accessed 9 December 2020.
18. Metcalf J, Crawford K. Where are human subjects in Big Data research? The emerging ethics divide. *Big Data Soc.* 2016;3(1):1–14.
19. Wang F, Casalino LP, Khullar D. Deep learning in medicine— Promise, progress, and challenges. *JAMA Intern Med*. 2019;179(3):293–294.
20. Lewis CM, Obregon-Tito A, Tito RY, et al. The human microbiome project: Lessons from human genomics. *Trends Microbiol*. 2012;20(1):1–4.
21. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias. 2016. ProPublica. 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Accessed 21 Dec 2020.
22. Massoglia M, Pridemore WA. Incarceration and health. *Annu Rev Sociol.* 2015;41(1):291–310.
23. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464): 447–453.
24. Heath S. Predictive Analytics Algorithm Displays Bias, Drives Inequity. 2019. Health IT Analytics. https://healthitanalytics.com/news/predictive-analytics-algorithm-displays-bias-drives-inequity. Accessed 9 December 2020.
25. Crawford K, Gray ML, Miltner K. Critiquing Big Data: politics, ethics, epistemology special section introduction. *Int J Commun Health*. 2014;8:1663–1672.
26. Cambridge Analytica controversy must spur researchers to update data ethics. *Nature.* 2018;555(7698):559–560.

27. McKay R. The right of privacy: emanations and intimations. *Mich Law Rev.* 1965;64(2):259–282.
28. Gavison, R. Privacy and the limits of law. *Yale Law J.* 1980;89(3):421–471
29. Aquisti A, Taylor C, Wagman L. The economics of privacy. *J Econ Lit.* 2016;54(2):442–492.
30. West S. Data capitalism: redefining the logistics of surveillance and privacy. *Bus Soc.* 2017;58(1):20–41.
31. Bonilla DN. Information management professionals working for intelligence organizations: ethics and deontology implications. *Secur Hum Rights.* 2014;24(3–4):264–279.
32. Markowitz A, Blaszkiewicz K, Montag C, et al. PsychoInformatics: Big Data shaping modern psychometrics. *Med Hypotheses.* 2014;82(4):405–411.
33. Harwell D. Wanted: The "Perfect Babysitter." Must Pass AI Scan for Respect and Attitude. 2018. The Washington Post. https://www.washingtonpost.com/technology/2018/11/16/wanted-perfect-babysitter-must-pass-ai-scan-respect-attitude/. Accessed 9 December 2020.
34. Youyou W, Kosinski M, Stillwell D. Computer-based personality judgments are more accurate than those made by humans. *Proc Natl Acad Sci USA.* 2015;112(4):1036–1040.
35. Molteni M. The U.S. Urgently Needs New Genetic Privacy Laws. 1 May 2019. https://www.wired.com/story/the-us-urgently-needs-new-genetic-privacy-laws/. Accessed 9 December 2020.
36. Erlich Y, Shor T, Pe'Er I, Carmi S. Identity inference of genomic data using long-range familial searches. *Science.* 2018;362(6415):690–694.
37. Glazer E, Tracey R, Horowitz J. FTC Approves Roughly $5 Billion Settlement with Facebook. 31 Mar 2020. The Wall Street Journal. https://www.wsj.com/articles/ftc-approves-roughly-5-billion-facebook-settlement-11562960538. Accessed 9 December 2020.
38. Dwoskin E. Zuckerberg Says He's Going All in On Private Messaging. Facebook's Declining User Numbers Tell Us Why. 2019. The Washington Post. https://www.washingtonpost.com/technology/2019/03/11/zuckerberg-says-hes-going-all-in-private-messaging-facebooks-declining-user-numbers-tell-us-why/.Accessed 9 December 2020.
39. Weisbaum H. Trust in Facebook Has Dropped by 66 Percent Since the Cambridge Analytica Scandal. 2018. NBC News. Available from: https://www.nbcnews.com/business/consumer/trust-facebookhas-dropped-51-percent-cambridge-analytica-scandal-n867011. Accessed 9 December 2020.
40. Mary, IRB. Focus Group at PRIM&R – PERVADE, 7 Nov 2018. https://pervade.umd.edu/2018/11/aerfocus/. Accessed 9 December 2020.
41. Shannon G, Jansen M, Williams K, et al. Gender equality on science, medicine, and global health: where are we at and why does it matter? *Lancet.* 2019;393(10171):560–569.
42. United States Department of Health and Human Services site. Covered Entities and Business Associates. Health Information Privacy. Available from: https://www.hhs.gov/hipaa/for-professionals/covered-entities/index.html. Accessed 9 December 2020.
43. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *N Engl J Med.* 2018;378(11):981–983.
44. Charles C, Gafnv A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes two to tango). *Soc Sci Med.* 1997;544(5):681–692.
45. Callebaut W. Scientific perspectivism: a philosopher of science's response to the challenge of Big Data biology. *Stud Hist Philos Sci A.* 2012;43(1):69–80.
46. Greene JA, Lea AS. Digital futures past—the long arc of Big Data in medicine. *N Engl J Med*. 2019;381(5):480–485.
47. Kuang C. Can A.I. Be Taught to Explain Itself? 2017. The New York Times Magazine. https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html. Accessed 9 December 2020.
48. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of Big Data in public health. *PLoS Comput Biol.* 2015;11(2):e1003904.
49. Ienca M, Ferretti A, Hurst S, et al. Considerations for ethics review of Big Data health research: a scoping review. *PLoS One.* 2018;13(10):e0204937.

ICNS