Governing AI-Driven Health Research: Are IRBs up to the task?

Friesen P, Douglas Jones R Marks, M, Pierce R, Fletcher K, Mishra A, Lorimer J, Veliz C, Hallowell N, Graham M, Chan MS, Davies H, Sallamuddin T

The Need for Ethics Oversight

AI tools such as machine learning, deep learning, and natural language processing are changing the nature of medical research. These technologies impact where health research is performed, who can conduct it, and which types of data are analyzed. As people increasingly record and upload details of their daily lives to the cloud, personal health data can now be inferred from non-medical data points, such as social media posts and browsing histories, or collected from consumer devices such as fitness trackers and smart speakers. From this data abundance, researchers build databases and train machine learning algorithms to predict, diagnose, and classify those whose data has been collected. This novel ability to create health data from unrelated online content has been called 'emergent medical data' and often takes place without the unawareness or consent of users [1]. Legal and regulatory systems have not caught up with these changes, leaving significant gaps in ethics oversight.

Outside of academia, where much of this medical research is taking place, the primary response to AI advances in medicine and other sectors has come in the form of AI ethics guidelines [2]. The recent proliferation of such guidelines reflects a shift in public perceptions of AI. News stories singing the praises of AI and forecasting its social benefits have given way to stories

documenting the misuse of consumer and patient data, and discussions of the threats AI poses to equality, democracy, and privacy. Growing anxieties have led critics to call for something stronger than mere ethical guidelines, demanding "governance with teeth" [3, 4]. In response, AI industry players have begun to develop their own variations on a familiar model of ethics oversight: the Institutional Review Board[1] (IRB). Versions of IRBs have been developed and implemented at Facebook and at Google's Deepmind, while nearly every significant technology company boasts some form of an ethics committee [5-7].

Within academic institutions, where IRBs have long played a role in research ethics oversight, responses have focused on adapting IRBs to capture the novel ethical issues posed by medical research involving AI. However, the model is being stretched, often unsuccessfully, to accommodate these forms of research [8]. In particular, AI research is posing significant challenges for IRBs that have developed along with a particular definition of 'human participants' and a narrow notion of the temporality of research and its risks.

Here, we examine the IRB as a model of ethics oversight for reviewing new forms of medical research involving AI, focusing on challenges that are being raised both in industry and in academia. As research collaborations across these two settings are becoming increasingly common, the significant differences in oversight between them are becoming more apparent. Because medical research involving AI often looks very similar across these two settings, however, we argue it is necessary to consider them in conjunction.

[1] Also known as the Research Ethics Committee (REC), Research Ethics Board (REB), Ethics Review Committee (ERC), and Ethics Committee (EC).

Given that AI medical research is often a global phenomenon, with risks that extend beyond borders, much of the discussion below is abstracted from particular regulatory structures. While we refer to IRBs throughout, a term common only in the United States, much of our analysis can be applied more widely, and we hope will be useful for a broader audience. In particular, we hope to initiate discussions amongst stakeholders who are invested in performing ethical review for AI-driven medical research, both in industry and academia, as well as those calling out for, and developing solutions of, more systemic oversight [9].

Below, we examine the origins of the IRB model and how medical research has changed drastically since the establishment of research ethics oversight. Next, using two case studies that blend the real and the hypothetical, we analyze the challenges IRBs are facing in both industry and institutional contexts as a result of AI-driven medical research. Finally, we argue that the model may be worth salvaging as a mechanism of oversight for health-related AI research and offer several recommendations of features IRBs will require if they are to prove effective in such a role.

Examining the Institutional Review Board Model

In the mid-twentieth century, widespread human rights violations occurred within a variety of research settings, from the Nazi concentration camps to the Tuskegee syphilis study. These events prompted the development of both research ethics standards, such as the Declaration of Helsinki and the Belmont Report, and a mechanism for applying those standards to research

proposals: the Institutional Review Board [10, 11]. IRBs consist of groups of experts and stakeholders (clinicians, scientists, community members) who review research protocols with an eye toward ethical concerns. They ensure that protocols comply with regulatory guidelines and have authority to withhold approval until such concerns have been addressed. This model of oversight was developed to prevent research abuses while preserving academic freedom for researchers and shielding research institutions from liability [12, 13].

When IRBs were developed, medical research looked very different than it does today. At the time, most health research was state-funded and performed by academic institutions. To conduct health research, investigators needed access to medical data, which could only be obtained in clinical settings. As a result, medical research was confined to those settings, and the resulting models for ethics oversight, including IRBs, were built around those constraints. Review requirements were grounded in features such as the receipt of government funding and links to an academic institution, while mechanisms to protect health information focused on medical data collected in clinical settings. Furthermore, when IRBs came into being, the paradigm of medical research was the clinical trial; research participants took part in experiments in person, and once a trial was over, no new risks were expected to appear. As a result, IRBs were designed to play an anticipatory role, predicting what risks might arise within research and ironing out ethical issues before they appeared [2]. Because the focus was on protecting individual research participants during ongoing trials, future harms and harms to communities were rarely taken into account [15].

[2] Still today, regulations in the United States explicitly instruct IRBs not to consider the "long-range effects of applying knowledge gained in the research" within the review process [14].

Today, medical research looks very different than it did back then. Machine learning techniques can be used to diagnose users on Twitter with post-partum depression or determine who has pancreatic cancer based on the symptoms they searched for on Google [16, 17]. Consequently, medical research is no longer restricted to those with access to medical records collected in clinical settings, although many national health privacy laws, such as HIPAA (Health Insurance Portability and Accountability Act), are restricted to such settings [18]. These changes have allowed medical research to be conducted outside of clinical and academic contexts by for-profit medical device, pharmaceutical, and internet companies, which often circumvent requirements for ethics oversight [19]. Furthermore, the changing nature of medical research means that much of it slips outside of the review process, even when such review is federally mandated. If data that is de-identified or publicly available is used, the research often is considered exempt from ethical review, regardless of whether risks related to re-identification, algorithmic bias, or data re-uses are present.

Falling Through the Cracks

The challenges that AI poses for IRBs are currently unfolding in different settings. In industry, where calls for ethics oversight are growing louder every day, IRBs are being turned to as a way to address gaps in oversight. However, in commercial contexts, they may be implemented in ways that serve corporate interests at the expense of ethics [20]. Meanwhile, academic and institutional IRBs, which have always grappled with ethical issues that arise with new forms of

research, are struggling to address those brought about by health research involving AI. Consider the following two examples:

Case 1: "Is it time for your organization to form an AI Ethics Committee?" [21]

> *Facebook plans to develop an algorithm to identify users at risk for suicide. The platform's AI will screen user posts using natural language processing; those classified as high risk will be sent to human moderators who will call the local authorities when they determine that an intervention is warranted [22]. Had this project been proposed within an academic setting, many ethical red flags would have been raised: the research involves sensitive mental health data, affects vulnerable populations, utilizes invasive interventions that the researchers have no control over, and users are not fully informed of the program or its risks. However, because Facebook falls outside the scope of regulations that require IRB review, the only ethics review that takes place involves a process developed and implemented by Facebook [7]. This review process resembles conventional IRBs in some ways: there are multiple levels of review, space for discretion, and it takes place before a project begins. However, it is conducted entirely by Facebook employees, and often only involves a direct line manager.*

Case 2: "Does this need ethics approval?"

> *An academic researcher plans to develop a facial recognition tool to identity individuals with fetal alcohol spectrum disorders (FASD). She hopes to collect publicly available*

images of individuals with FASD from across the web (e.g. educational sites, social media sites), mix these photos with images of individuals without FASD, and use this data set to train an algorithm to identify individuals with FASD. She isn't sure if she needs ethics approval before she begins the project. Consulting a decision tree on her university's website, she considers whether her research involves "data/specimens about or from individuals who are or may be still living" [23]. The research will involve data from individuals who are still alive. She consults the decision tree again, considering if "all the information about the specimens/data available in the public", which it is. This points her towards her answer: "Project is not human subjects research". No application or ethics review is required, so she proceeds with the research. Once it has been published, the classification model she developed is adopted by a schoolboard in order to identify children with FASD early, so teachers can be alerted as to which students may lack impulse control.

These cases highlight current challenges arising at the intersection of IRBs and health research involving AI. In the first case, which looks at Facebook's suicide prevention efforts and its development of an internal IRB (both components of this case are real but their conjunction is hypothetical), ethics review does takes place. However, the *kind* of review and the *kind* of committee are ethically suspect: the IRB was developed and implemented by Facebook, and other company priorities may have influenced the review process. Furthermore, a lack of transparency regarding the process, and the individuals involved, ensures that no one outside the company can assess the project's safety and efficacy. This opacity raises concerns related to the authority and independence of Facebook's IRB and exemplifies the significant risk that

organizations will adopt IRBs as a form of "ethics washing", in which ethics committees become a cynical public relations exercise [24].

The second case, which is hypothetical but involves current review procedures from a university in the United States, highlights the importance of revising the boundaries of what is considered health research and updating processes of research ethics oversight to account for varying data lifecycles. In this case, ethics review does not take place since the proposal does not involve research participants in the traditional sense. Since IRBs were developed with flesh and blood participants in mind, participants like the ones in this example don't trigger review processes. While images of their faces were used to develop facial recognition technologies, which enabled the development of a new classification system, they are not considered research participants because their data is publicly available online. This example suggests that our processes for delineating health research in need of ethics review requires revision. Furthermore, even if review had been triggered, the review process would not, at present, explicitly ask about data sources, data lifecycles, or the potential re-use or misuse of research outputs, even though the models developed can be used for purposes beyond those originally intended, and can put particular, often marginalized, populations at risk. This example highlights the shortcomings of relying solely on anticipatory review for medical research involving new technologies. Data can be recycled and re-used in unanticipated contexts, and unless IRBs are updated to reflect these trends, society risks repeating human rights violations like those that prompted the development of IRBs in the mid-twentieth century.

While many of the challenges IRBs face are not unique to AI health research (e.g., 'research participant' has always been hard to define), the rate of growth and ethical impact of this field requires reflection on the promise and limits of the IRB model of oversight.

Adapting Institutional Review Boards for AI

While these examples may seem to suggest that IRBs are either the newest fad at the ethics launderette or an outdated model of oversight, we contend that IRBs are worth adapting as a mechanism of oversight for AI health research. Crucially, the deliberative process central to IRBs, in which space is given to reflecting on the risks and benefits that might be involved in novel research projects, is well suited to the complexity and unpredictability of AI health research. Furthermore, bringing together diverse stakeholders and experts is worthwhile when the impact of research can be significant, difficult to foresee, and unlikely to be understood by any single expert, as is the case with AI-driven medical research.

We acknowledge, however, that IRBs currently face a variety of issues. These include a lack of consistency in decision-making within and across committees [25-27], a lack of transparency [28], poor representation of the researchers and publics they are meant to represent [29, 30], insufficient training [31], and a lack of measures to examine their effectiveness [32]. We also acknowledge that questions related to enforcement loom large in the background of these recommendations and remain unanswered. While expanding and adapting regulations will look different in various legislative contexts, in many settings it will require rethinking traditional triggers for research ethics review, including the thresholds for determining what constitutes

'research' and 'human subjects', what data is considered to be 'medical', as well as the weight we place on 'reasonable expectations of privacy' and that which is 'publicly available'.

With these limitations in mind, we offer several recommendations for how IRBs might be extended and adapted in order to better accommodate medical research that utilizes AI tools and techniques. In particular, we identify eight features that we see as crucial to the success of IRBs as a form of oversight in the domain of AI-driven health research. Some of these features are already present in institutional IRBs, at least in principle, while many of them are lacking in ethics committees that have been developed within industry.

Despite their differences, these recommendations apply to both institutional and commercial ethics committees. Medical research taking place across these two settings is often strikingly similar, involving similar risks and similar data subjects. Regardless of whether their data is being collected and examined by a university or corporate researcher, these data subjects have legitimate interests in how their data is used and these interests merit protection in the form of oversight. However, as described above, the review process can look drastically different in each setting, and in both cases, significant gaps exist. Given the frequency with which medical data is now being created about unsuspecting users without their awareness or consent in both of these domains, it is no longer acceptable for the review process across academic and commercial IRBs to differ so greatly. Thus, we consider them in tandem.

1. Independence

It is crucial to ensure the independence of IRBs in relation to the research projects they oversee. IRB members ought to be distinct from the research team and those who provide its funding. This distance can reduce bias and make it more likely that members feel comfortable criticizing a project without fear of repercussions. Independence is particularly important in industry-led projects, where shifting ethics oversight out of the hands of managers and CEOs ensures that one's analysis of ethical issues is not shaped by corporate goals.

2. Authority

IRBs must be given authority to require changes to research protocols before research can begin. A recent industry trend involves the development of ethics committees that serve in an advisory capacity, but such committees leave the decision-making authority in the hands of industry players[3]. Advice is not the same as oversight, and it can be ignored if it is incompatible with a company's economic goals.

3. Transparency

The process of drawing a line between what is considered permissible and impermissible in AI-driven health research should be reasoned, justified, and published. This will allow for researchers, IRBs, and outside observers to learn from each other and for the review process to improve over time. While transparency in industry-based committees raises unique challenges related to intellectual property and responsibilities to shareholders, there are tools available,

[3] An interesting exception is Axon's ethics board (see [33]).

including pseudo-code, non-disclosure agreements, and pre-competitive alliances that can help overcome these constraints.

4. Diversity

Representation has long been a central issue related to IRBs, and it is even more vital in medical AI projects involving vulnerable participants with limited consent processes (e.g. individuals at risk of suicide). To reduce the likelihood of bias and exploitation, it is essential to include representatives of the populations most likely to be affected by AI health research. Involving experts who have studied past discrimination against marginalized communities may also reduce the adverse effects of such research [34].

5. Proportional Review

The presence and degree of ethics review should be proportional to the risk and uncertainty present within the research (see Table 1 for suggestions regarding how a triage process could work), and not dependent on who is funding the research or whether it takes place in a commercial context. Risk here should not be understood solely as the risk to each research participant; it should include the risks of the research project as a whole, including potential secondary uses of the data or technology. An overextension of ethics oversight should also be avoided. Low-risk projects could be adequately self-reviewed (e.g., using Data Protection Impact Assessments to demonstrate GDPR (General Data Protection Regulation) compliance [35]), while high risk projects with uncertain impacts may require more extensive review. For novel

forms of research that raise unique ethical issues, oversight by national, international, or institutional committees may be ideal (e.g., as with Embryonic Stem Cell Research Committees [36]).

6. Ongoing Monitoring

AI-driven health research disrupts previous conceptions of research, in which there was a clear beginning and an end, and the participants involved were easy to identify [37]. Now, data lifecycles vary greatly, and models developed in research contexts can later be applied to new data, creating novel health data about unsuspecting users. As a result, anticipatory review from IRBs may not be sufficient, and in some cases, ongoing monitoring may be required. Such monitoring could be overseen by IRBs or by distinct committees, perhaps modelled after Data Safety Monitoring Boards (DSMBs) [4].

7. Standards and Tools

Measures must be taken to develop consistency within and across IRB procedures. While variation is to some extent inevitable (and in some cases desirable), consistency can be improved through clear procedural guidelines relevant to a research context [25]. Tools can be used to help committee members and investigators engage in ethical analysis and forecasting, many of which

[4] Another solution proposed to counteract these worries is to develop AI licenses, which introduce restrictions on the use, reproduction, and distribution of software [38].

have already been developed within the space of AI research (e.g. Algorithmic Impact Assessments) [39, 40].

8. Evidence of Effectiveness

Finally, IRBs currently suffer from a lack of evaluative processes, leaving us with a dearth of evidence with regards to whether they are effective. While developing such processes is no easy task, given the number of goals IRBs juggle (protecting participants, promoting research, ensuring justice within research), dedicated efforts to measure IRB effectiveness should be made within the domain of AI health research [41].

Looking Forward

The Institutional Review Board is an established model of research ethics oversight. However, the increase in AI-driven research practices in medicine pose challenges for both committees that already exist and those in formation. While crucial questions remain surrounding how, and at what levels, enforcement should occur, we hope the points made above can inform emerging conversations among stakeholders regarding the development and adaptation of IRBs in industry, academia, and collaborations between the two. In order to ensure that adequate processes of oversight exist for AI health research, regardless of where it takes place, IRBs in both settings will need to evolve. Institutional ethics committees must adapt their scope and discussions to account for new versions of research that are becoming commonplace, in which data lifecycles are no longer predictable, consent is often overlooked, and the downstream implications of

research are not always visible at the outset. For newly established IRBs in industry, retaining features such as independence, transparency, and authority will be vital. Since ethics committees are increasingly turned to in order to fill gaps in ethics oversight of health-related AI research, it is essential that we attend to the ways the reinvention of research is reinventing the ways it must be governed.

References

1. Marks, M., *Emergent Medical Data*, in *Bill of Health*. 2017.
2. Algorithm Watch. *AI Ethics Guidelines Global Inventory*. 2019 [cited 2019 June 27]; Available from: https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/.
3. Article 19, *Governance with teeth: How human rights can strengthen FAT and ethics initiatives on artificial intelligence*. 2019.
4. Whittaker, M., et al., *AI now report 2018*. 2018: AI Now Institute at New York University.
5. Sandler, R. and J. Basl, *Building data and AI ethics committees*. 2019, Accenture.
6. Tiell, S., *Create an Ethics Committee to Keep Your AI Initiative in Check*, in *Harvard Business Review*. 2019.
7. Jackman, M. and L. Kanerva, *Evolving the IRB: Building robust review for industry research.* Washington and Lee Law Review Online, 2016. **72**(3): p. 442.
8. Samuel, G., G.E. Derrick, and T. van Leeuwen, *The ethics ecosystem: Personal ethics, network governance and regulating actors governing the use of social media research data.* Minerva, 2019. **57**(3): p. 317-343.
9. Vayena, E. and A. Blasimme, *Health Research with Big Data: Time for Systemic Oversight.* The Journal of Law, Medicine & Ethics, 2018. **46**(1): p. 119-129.
10. *World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects.* J Am Coll Dent, 2014. **81**(3): p. 14-8.
11. Ryan, K., et al., *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research.*, U.S.N.C.f.t.P.o.H.S.o.B. and and B. Research, Editors. 1979: Washington, D.C.
12. Stark, L., *Behind closed doors: IRBs and the making of ethical research*. 2011: University of Chicago Press.
13. Hedgecoe, A., *Reputational Risk, Academic Freedom and Research Ethics Review.* Sociology, 2015. **50**(3): p. 486-501.
14. Resnik, D.B., *Dual-use review and the IRB.* Journal of clinical research best practices, 2010. **6**(1).
15. Weijer, C., *Protecting communities in research: philosophical and pragmatic challenges.* Camb Q Healthc Ethics, 1999. **8**(4): p. 501-13.
16. Paparrizos, J., R.W. White, and E. Horvitz, *Screening for Pancreatic Adenocarcinoma Using Signals From Web Search Logs: Feasibility Study and Results.* Journal of Oncology Practice, 2016.
17. De Choudhury, M., S. Counts, and E. Horvitz. *Predicting postpartum changes in emotion and behavior via social media*. in *Proceedings of the SIGCHI conference on human factors in computing systems*. 2013. ACM.
18. Marks, M., *Tech companies dangerous practice: Using AI to infer hidden health data*, in *STAT News*. 2019.
19. Stoeklé, H.-C., et al., *23andMe: a new two-sided data-banking market model.* BMC Med Ethics, 2016. **17**: p. 19-19.
20. Metcalf, J. and E. Moss, *Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics.* Social Research: An International Quarterly, 2019. **86**(2): p. 449-476.

21. Morphy, E., *Is It Time for Your Organization to Form an AI Ethics Committee?*, in *CMS Wire*. 2019.
22. Marks, M., *Artificial Intelligence Based Suicide Prediction.* The Yale Journal of Health Policy, Law, and Ethics, 2019((forthcoming)).
23. New York University. *IRB Decision Tree*. 2019 Nov 29, 2019]; Available from: https://www.nyu.edu/content/dam/nyu/research/documents/IRB/IRBDecisionTree.pdf.
24. Wagner, B., *Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?* Being Profiling. Cogitas Ergo Sum, 2018.
25. Friesen, P., A. Yusof, and M. Sheehan, *Should the Decisions of Institutional Review Boards be Consistent?* Ethics and Human Research, 2019.
26. Angell, E.L., et al., *Is' inconsistency'in research ethics committee decision-making really a problem? An empirical investigation and reflection.* Clinical Ethics, 2007. **2**(2): p. 92-99.
27. Khan, M.A., et al., *Variability of the Institutional Review Board Process Within a National Research Network.* Clinical Pediatrics, 2014. **53**(6): p. 556-560.
28. Lynch, H.F., *Opening Closed Doors: Promoting IRB Transparency.* The Journal of Law, Medicine & Ethics, 2018. **46**(1): p. 145-158.
29. De Vries, R., D.A. DeBruin, and A. Goodgame, *Ethics review of social, behavioral, and economic research: Where should we go from here?* Ethics & behavior, 2004. **14**(4): p. 351-368.
30. Keith-Spiegel, P., G. Koocher, and B. Tabachnick, *What scientists want from their research ethics committee.* Journal of empirical research on human research ethics : JERHRE, 2006. **1**(1): p. 67-82.
31. Abbott, L. and C. Grady, *A systematic review of the empirical literature evaluating IRBs: what we know and what we still need to learn.* Journal of Empirical Research on Human Research Ethics, 2011. **6**(1): p. 3-19.
32. Grady, C., *Institutional review boards: Purpose and challenges.* CHEST Journal, 2015. **148**(5): p. 1148-1155.
33. Warzel, C., *A Major Police Body Cam Company Just Banned Facial Recognition*, in *The New York Times*. 2019.
34. Nordling, L., *Mind the Gap*, in *Nature*. 2019. p. S103 - S105.
35. *Data Protection Impact Assessment (DPIA)*. 2019 Nov 29, 2019]; Available from: https://gdpr.eu/data-protection-impact-assessment-template/.
36. Friesen, P., B. Redman, and A. Caplan, *Of Straws, Camels, Research Regulation, and IRBs.* Therapeutic Innovation & Regulatory Science, 2018. **53**(4): p. 526-534.
37. Samuel, G., et al., *Is It Time to Re-Evaluate the Ethics Governance of Social Media Research?* Journal of Empirical Research on Human Research Ethics, 2018. **13**(4): p. 452-454.
38. *Responsible AI Licenses*. [cited 2019 November 22]; Available from: https://www.licenses.ai/.
39. Ballard, S. and R. Calo, *Taking Futures Seriously: Forecasting as Method in Robotics Law and Policy*, in *We Robot 2019*. 2019.
40. Reisman, D., et al., *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability* 2018.

41. Lynch, H.F., et al., *Of parachutes and participant protection: Moving beyond quality to advance effective research ethics oversight.* Journal of Empirical Research on Human Research Ethics, 2018: p. 1556264618812625.

Table 1.

Features to guide triage for health research involving AI:

- potential for re-identification
- degree of explainability
- sensitivity of data
- expectations of users
- potential for dual use or discrimination
- risk of community or population level harms
- vulnerabilities in participants or tools
- the presence and method of informed consent
- connections to other crucial systems
- likelihood that significant or sensitive inferences will be drawn from the results