

**Gabrielle Samuel & Gemma Derrick**

## **Ethical standards in population health research**

**This online first version has been peer-reviewed, accepted and edited,  
but not formatted and finalized with corrections from authors and proofreaders.**

# **Defining ethical standards for the application of digital tools to population health research**

Gabrielle Samuel<sup>a</sup> & Gemma Derrick<sup>b</sup>

<sup>a</sup> Department of Global Health and Social Medicine, King's College London, Bush House, North East Wing, The Strand, London, WC2R 2LS, England.

<sup>b</sup> Department of Educational Research, Lancaster University, Lancaster, England.

Correspondence to Gabrielle Samuel (email: [gabrielle.samuel@kcl.ac.uk](mailto:gabrielle.samuel@kcl.ac.uk)).

(Submitted: 15 May 2019 – Revised version received: 18 October 2019 – Accepted: 25 October 2019 – Published online: 17 January 2020)

### **Abstract**

There is growing interest in population health research which uses methods based on artificial intelligence. Such research draws on a range of clinical and non-clinical data to make predictions about health risks, such as identifying epidemics and monitoring disease spread. Much of this research uses data from social media in the public domain or anonymous secondary health data and is therefore exempt from ethics committee scrutiny. While the ethical use and regulation of digital-based research has been discussed, little attention has been given to the ethics governance of such research in higher education institutions in the field of population health. Such governance is essential to how scholars make ethical decisions and provides assurance to the public that researchers are acting ethically. We propose a process of ethics governance for population health research in higher education institutions. The approach takes the form of review after the research has been completed, with particular focus on the role artificial intelligence algorithms play in augmenting decision-making. The first layer of review could be national, open-science repositories for open-source algorithms and affiliated data or information which are developed during research. The second layer would be a sector-specific validation of the research processes and algorithms by a committee of academics and stakeholders with a wide range of expertise across disciplines. The committee could be created as an off-shoot of an already functioning national oversight body or health technology assessment organization. We use case studies of good practice to explore how this process might operate.

### **Introduction**

The co-founder of the independent thinktank DataEthics has observed that “data ethics is [now] fashionable.”<sup>1</sup> This is true: international attention in both the private and public sector is becoming focused on the ethical implications of using digital data and artificial intelligence in both the health and non-health arena. Such scrutiny includes questions around data governance, data minimization (only collecting and processing data which are needed for

specific purposes) and protecting the privacy of data subjects; considerations of consent and trust; concerns associated with data accountability, transparency and explainability; as well as issues related to fairness, justice and bias in data sets. It is concerning, however, that discussions are lacking about how to update ethics governance in higher education institutions to move towards a shared understanding of ethics best practice for research that uses digital data.<sup>2</sup> In the public sector, ethics governance in higher education is central to the way scholars make decisions about ethical value, and frames the norms of what becomes considered ethically acceptable by the whole research community. Ethics governance also assures the public that publicly-funded researchers are acting ethically and that higher education institutions can be trusted. These assurances are even more important in the realm of public health research, where the risk to society – in terms of predictions which could lead to health over-surveillance or inequity – are particularly high.

Ethical behaviour in higher education research means ensuring adherence to shared understandings of appropriate researcher practices throughout the research process from inception to dissemination and beyond. We have argued previously that such behaviour is reinforced through an informal, research-governed process which we termed an ethics ecosystem.<sup>3</sup> An ethics ecosystem is an interconnected network of researchers, research institutions and external bodies (publishers, funding bodies, professional associations and their policies) who participate equally in the promotion, evaluation and enforcement of ethically responsible research behaviour. However, the development of a stable ethics ecosystem in higher education institutions has become increasingly challenging in the areas of research which use innovative methods such as artificial intelligence. This is because questions remain about how to manage, process and interpret data predictions in an ethically responsible manner. As such, there are no shared norms and understandings of how to conduct such research ethically; traditional tools for ensuring ethical research behaviour that focus solely on consent and privacy are losing relevance because they are insufficient to deal with the range of ethical issues raised by digital research. As researchers strive to reach new shared understandings of ethical practice, a culture of personal ethics has emerged whereby researchers monitor their own (often different) decisions about how best to act ethically, without being subject to accountability or audit by other parts of the ethics ecosystem.<sup>3</sup>

Normally, research ethics committees in higher education institutions are tasked with the role of ensuring ethics best practice for all research involving human participants. However, at least in the United Kingdom of Great Britain and Northern Ireland, a large

proportion of health research which uses artificial intelligence-based methods to analyse data is exempt from ethics committee scrutiny (Samuel G, Department of Global Health and Social Medicine, King's College London, England, unpublished observations, 2019). This practice is particularly true for research methods which include the use of, for example, social media data in the public domain, geolocation data and anonymous secondary health data for which a licensing agreement has been signed. This lack of ethics scrutiny is compounded by inconsistent ethics guidelines for publishing such research in international peer-reviewed journals.<sup>3</sup>

There is growing interest in population health research which uses artificial intelligence-based methods. Such research draws on a range of clinical and non-clinical data to make predictions about health risks, such as identifying epidemics and monitoring disease spread. Any public health tool that uses such data could potentially be developed without systematic ethics oversight by a higher education institution, and (potentially) little reflection on the public interest or concomitant risks attached to making health predictions from the tool. Even when there is oversight from a research ethics committee, we observe that committee members often lack the experience or confidence regarding particular issues associated with digital research (Samuel G, Department of Global Health and Social Medicine, King's College London, England, unpublished observations, 2019).<sup>4</sup> To address this gap, we propose a model for ethical scrutiny of artificial intelligence-based research in public health.

## Approaches to ethics oversight

Commentators are only recently beginning to assert the need for more agreed-upon frameworks for ethics governance in higher education institutions.<sup>5</sup> Awareness of the issues has been driven to some extent by a recent international, high-profile example of inappropriate ethical behaviour by a researcher in a United Kingdom higher education institution. The researcher designed software which was used by a consulting firm to access personal data from millions of social media users without their consent and which was used for political advertising purposes.<sup>6</sup> Questions remain, however, regarding what a framework for ethics governance would look like in practice, and how, in some jurisdictions, such a system would guard against ethics dumping (exporting unethical research practices, for example, unethical data processing) to countries where research ethics committee oversight is lacking.

An approach to ethics oversight, which we focus on here, could be an extra layer of governance in the ethics ecosystem after the research has been conducted, that is, an ex-post review. This approach would be particularly important when the research involves artificial intelligence-based algorithms which have a role in augmented decision-making. This is because many of the ethical concerns voiced about these technologies relate to their impact on society rather than to questions around the research itself. As we discuss below, the system would be analogous to premarket approval in drug regulation. Just as drug regulation aims to protect public health by ensuring the safety and efficacy of drugs, so too could ex-post review mitigate any societal harm potentially arising from artificial intelligence-based algorithms designed to make predictions about health. The review process would safeguard public health by minimizing the risk of harm caused, for example, by placing an over-reliance on artificial intelligence for decision-making in cases where it could potentially make inaccurate or unfair predictions. Such instances of harm have been reported in other sectors, and in the health sector the failure of Google Flu Trends reminds us to be cautious about the claims made for artificial intelligence.<sup>7,8</sup> An algorithm was developed by Google to predict influenza outbreaks from the analysis of people's searches for information assumed to relate to influenza on Google's online search engine. The algorithm missed the peak of the 2013 influenza season in the United States of America because the search terms used in the construction of the algorithm produced inaccurate prediction results.<sup>8</sup>

The subject of ex-post review for oversight of the use of artificial intelligence in research is being discussed in the literature.<sup>9–11</sup> Commentators have called for an artificial intelligence ombudsperson to ensure the auditing of allegedly unfair or inequitable uses of artificial intelligence. One proposed method is the adoption of trust labels – described as labels which certify the trustworthiness of the algorithm – to ensure people understand the merits of artificial intelligence-based methods.<sup>10</sup> Others talk about a potential licensing system to ensure quality control, analogous to licensing in many other sectors, such as in production and manufacturing.<sup>11</sup>

There have been doubts that any ex-post governance system can overcome the many obstacles required to cover the multiple evolving fields of digital research more generally.<sup>11</sup> Nevertheless, a sector-specific<sup>12</sup> and discipline-specific approach, particularly in health research – which is already well accustomed to such oversight and regulation – seems feasible at a national level. As higher education institutions have an emerging role in promoting their social role, oversight of artificial intelligence-based research is in line with

recent principles underlying the practice of responsible research and innovation.<sup>13</sup> We therefore believe that researchers should not retreat from the responsibility of ensuring such a system is established as part of the ethics ecosystem of higher education institutions.

## Proposed review model

The ex-post review model we propose is not a finished product but a starting point to drive discussion within national public health and broader research communities towards committing to such an approach or a similar one. The model is two-layered, with ex-post review working at the second layer, though the system could potentially work with only the second layer.

The first layer of review would require a systematic, open-science infrastructure for centralized national repositories for open-source algorithms and affiliated data or information which are developed during the research process and which are intended for wide use beyond the research studies, particularly for decision-making. These repositories could be health-specific or non-health-specific, though the latter would be more feasible because the distinction between health and non-health research can be unclear and is not well defined within higher education institutions. A culture of open science and sharing of innovation is already promoted nationally and internationally,<sup>14</sup> and expanding its remit for algorithms and associated workflows and software is not a daunting task.

Any open-science repository needs to be managed, curated and driven by higher education institutions and funding bodies, with clear incentives for compliance. For example, to develop best practice, there should be a requirement for algorithms and affiliated data to be placed in repositories just as there is for research data in many jurisdictions. We assume that in some instances access to data will need to be restricted to certain stakeholders, for certain data sets and certain circumstances. While this first layer is not a requirement for the second ex-post review layer, we believe that there are advantages to making artificial intelligence algorithms and the affiliated data more accessible. An open-science repository creates a way for other researchers and stakeholders to test the algorithms with their own data, checking for spurious predictions and highlighting any concerns or issues which may be present within the artificial-intelligence prediction models. This open culture will have the added effect of driving innovation, because models can then be corrected and built on to achieve better predictions for the future.

The second layer of the ex-post review process would be a sector-specific validation of the research processes and algorithms. A sector-specific approach has also been suggested elsewhere,<sup>15</sup> and makes sense because of its feasibility and ability to accommodate the specific needs of the health sector. These benefits mean that the review process can overcome some of the doubts raised about an ex-post governance system more generally, which would be difficult to implement across different sectors. A sector-specific approach also makes the sense because regulatory systems are typically sector-specific. Consider the use of DNA (deoxyribonucleic acid) testing, for example, which is regulated differently in the criminal justice system than within health care and even wider society. Ex-post review would work best if it included the products of research not only from higher education institutions but also private sector organizations. We propose such inclusion because much of the research we refer to is being carried out in the private sector and artificial intelligence algorithms developed in academia often require private-sector investment to make them feasible for wider public use.

## Suggestions for implementation

Analogous systems of ex-post review of innovative prediction algorithms already exist in other sectors. In the Netherlands, for example, ex-post review is a legal requirement for the use of forensic DNA phenotyping in the criminal justice system. At present, if law enforcement officers have DNA samples of unknown origin from a crime scene, they can have the DNA tested (under specific circumstances) in such a way that makes predictions about certain externally visible characteristics of the person from whom the DNA originated. Legislation permits law enforcement officers to use only those tests which have been validated and reviewed and, at present, the law covers phenotype testing of hair and eye colour. For any new phenotype testing model to be considered for use, the system requires researchers to publish their validated models openly as a series of papers (similar to layer 1 of our model above). These papers are then submitted to the parliament of the Netherlands for review and for checking whether the models have been validated appropriately for use within the criminal justice system. Parliament therefore functions as an ombudsperson to provide sector-specific ex-post review of the testing models to ensure the underlying science is rigorous and validated.

Within the sphere of public health, ex-post review could be conducted by a committee comprising academics and stakeholders (such as lay people, professionals or users of the technology) with a wide range of expertise across disciplines, including but not limited to

health, medicine, artificial intelligence research, social science and ethics. The aim of the committee would be to mitigate the risks of potential harm which could be caused by the technology as much as possible by: reviewing scientific questions relating to the origin and quality of the data, algorithms and artificial intelligence; confirming the validation steps which have been conducted to ensure the prediction models work; and requesting further validation to be carried out, if required. This risk assessment could progress faster when researchers have already performed their own assessment of the societal impact of their research. As suggested elsewhere,<sup>16</sup> this approach could also consider the integrity of participants' data used in the research, and broader questions of social justice and value as they apply to the particular national jurisdiction. The committee could simply review or validate the research, have a more regulatory role or oversee any trialling of a research tool; in this way it would act as the artificial intelligence ombudsperson as discussed above. The committee may also be required to have a dynamic regulatory role, especially for algorithms whose performance evolves as more data are added to their systems. However, questions around the parameters of how this regulatory role could be put into practice still need answering – one suggestion being provisional licensing. One example of such an approach in the making is the International Telecommunication Union and World Health Organization focus group on artificial intelligence and health,<sup>17</sup> which is working towards collecting a repository of data to test artificial intelligence technology for health within a standardized benchmarking framework.

At a national level, the establishment of a special committee might be burdensome in terms of slowing down the research process because it requires an additional element in the ethics ecosystem. Nevertheless, we argue that it is important to have a system of ethics governance when the research has a potentially high impact on the health of society. The costs could be minimized by streamlining the process. To add capacity, for example, the committee could be created as an off-shoot of an already functioning national oversight body or health technology assessment organization. Which organization would depend on the committees' exact function in each jurisdiction. In the United Kingdom, for example, suitable candidates include the National Institute for Health and Care Excellence, the Medicines and Health care products Regulatory Agency or the newly founded Centre for Data Ethics and Innovation. All of these organizations are in the process of producing guidelines about the responsible use of artificial intelligence in health care. In fact, the United States Food and Drug Administration and the European Medicines Agency, which already regulate drug use

in their respective territories via ex-post review, are being called upon to fulfil such a role for overseeing medical devices that use artificial intelligence in health care.<sup>18</sup> Adding an additional division to focus specifically on the ex-post review of artificial intelligence-based research in the domain of public health would therefore be logical. Existing regulatory agencies are likely to be the most suitable candidates for such a role, and the European and United Kingdom's agencies have now started introducing measures to scrutinize medical software. The measures include clarifying guidelines for medical devices in the light of digital data and software, and providing recommendations for the standardization of artificial intelligence in the medical device health-care sector. The Food and Drug Administration has now approved the first use of artificial intelligence to diagnose eye disease.<sup>19</sup> Unclear boundaries over what constitutes a medical device, however, may leave some population-health specific algorithms outside of the remit of these organizations. Moreover, algorithms are context- and population-specific and will need validation in each national jurisdiction.

Together, national infrastructures which incorporate ex-post review for artificial intelligence-based algorithms designed for health applications can start to re-balance the ethics ecosystem. In this way nations would begin the process of developing a new shared understanding of ethics best practice for artificial intelligence-based public health research. Within this best practice, artificial intelligence-associated research will be openly scrutinized before any application that affects wider society is disseminated and used, to minimize as much as possible the potential for these systems to cause harm. Whether artificial intelligence-based health research continues to generate new ethical concerns or becomes just one more method in a researcher's toolbox we must take care to avoid any unintended consequences of big data studies.

**Funding:**

The authors received a Seed Award in Humanities and Social Science from the Wellcome Trust for the project entitled "The ethical governance of artificial intelligence health research in higher education institutions," grant number: 213619/Z/18/Z/.

**Competing interests:**

None declared.



## References

1. Hasselbalch G. Data ethics is a game of interests [internet]. Copenhagen: DataEthics; 2018. Available from: <https://dataethics.eu/data-ethics-is-a-game-of-interests/> [cited 2019 Dec 4].
2. Gasser U. The ethics and governance of artificial intelligence: on the role of universities [internet]. Boston: Berkman Klein Center for Internet and Society at Harvard University; 2017. Available from: <https://medium.com/berkman-klein-center/the-ethics-and-governance-of-ai-on-the-role-of-universities-6c31393fe602> [cited 2019 Dec 4].
3. Samuel G, Derrick GE, van Leeuwen T. The ethics ecosystem: personal ethics, network governance and regulating actors governing the use of social media research data. *Minerva*. 2019 Sep;57(3):317–43. <https://doi.org/10.1007/s11024-019-09368-3> PMID:31501635
4. Sellers C, Samuel G, Derrick G. Reasoning “uncharted territory”: notions of expertise within ethics review panels assessing research use of social media. *J Empir Res Hum Res Ethics*. 2019 Dec 12;51:1556264619837088. <https://doi.org/10.1177/1556264619837088> PMID:31826689
5. Raymond N. Safeguards for human studies can’t cope with big data. *Nature*. 2019 04;568(7752):277. <https://doi.org/10.1038/d41586-019-01164-z> PMID:30988466
6. Isaak J, Hanna MJ. User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer*. 2018 Aug;51(8):56–9. <https://doi.org/10.1109/MC.2018.3191268>
7. Corbett-Davies S, Pierson E, Feller A, Goel S. A computer program used for bail and sentencing decisions was labeled biased against blacks. It’s actually not that clear. *Washington Post*. 2016 Oct 17. Available from: <https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/> [cited 2019 Dec 4].
8. Lazer D, Kennedy R. What we can learn from the epic failure of Google Flu trends [internet]. *Wired Magazine*. 2015 Oct 1. Available from: <https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/> [cited 2019 Dec 4].
9. Cath C, Zimmer M, Lomborg S, Zevenbergen B. Association of Internet Researchers (AoIR) roundtable summary: artificial intelligence and the good society workshop proceedings. *Philos Technol*. 2018;31(1):155–62. <https://doi.org/10.1007/s13347-018-0304-8>
10. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People – an ethical framework for a good artificial intelligence society: opportunities, risks, principles, and recommendations. *Minds Mach (Dordr)*. 2018;28(4):689–707. <https://doi.org/10.1007/s11023-018-9482-5> PMID:30930541
11. Algorithms and human rights: study on the human rights dimensions of automated data processing techniques and possible regulatory implications. Council of Europe Study DGI(2017)12. Strasbourg; Committee of Experts on

Internet Intermediaries, Council of Europe; 2018. Available from: <https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html> [cited 2019 Dec 4].

12. Whittaker M, Crawford K, Dobbe R, Fried G, Kaziunas E, Mathur V, et al. AI Now report. New York: AI Now Institute at New York University; 2018. [cited 2019 Dec 4]. Available from: [https://ainowinstitute.org/AI\\_Now\\_2018\\_Report.pdf](https://ainowinstitute.org/AI_Now_2018_Report.pdf)
13. Marginson S. Higher education and public good. *High Educ Q.* 2011;65(4):411–33. <https://doi.org/10.1111/j.1468-2273.2011.00496.x>
14. Digital health atlas [internet]. Geneva: World Health Organization; 2019. Available from: <https://digitalhealthatlas.org/en/-/> [cited 2019 Dec 4].
15. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. *NPJ Digit Med.* 2019 05 13;2(1):38. <https://doi.org/10.1038/s41746-019-0111-3> PMID:31304384
16. Price WN. Artificial intelligence in health care: applications and legal implications [dissertation]. Ann Arbor: University of Michigan Law School; 2017. Available from: <https://repository.law.umich.edu/cgi/viewcontent.cgi?article=2932&context=articles> [cited 2019 Dec 4].
17. Focus Group on Artificial Intelligence for Health [internet]. Geneva: International Telecommunication Union; 2019. Available from: <https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx> [cited 2019 Dec 4].
18. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD). Discussion paper and request for feedback. Silver Spring: United States Food and Drug Administration; 2019. Available from: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device> [cited 2019 Dec 4].
19. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. Silver Spring: United States Food and Drug Administration; 2018. Available from: <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye> [cited 2019 Dec 4].