AUTHORS:
Sharon Kling[1,2]
Shenuka Singh[3]
Theresa L. Burgess[1,4]
Gonasagrie Nair[1]

AFFILIATIONS:
[1]Centre for Medical Ethics and Law, WHO Bioethics Collaborating Centre, Department of Medicine, Stellenbosch University, Cape Town, South Africa
[2]Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa
[3]Discipline of Dentistry, School of Health Sciences, University of KwaZulu-Natal, Durban, South Africa
[4]Division of Physiotherapy, Department of Health and Rehabilitation Sciences, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

CORRESPONDENCE TO:
Sharon Kling

EMAIL:
sk@sun.ac.za

HOW TO CITE:
Kling S, Singh S, Burgess TL, Nair G. The role of an ethics advisory committee in data science research in sub-Saharan Africa. S Afr J Sci. 2023;119(5/6), Art. #14724. https://doi.org/10.17159/sajs.2023/14724

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
ethics, committee, data science, research, Africa

PUBLISHED:
30 May 2023

The role of an ethics advisory committee in data science research in sub-Saharan Africa

Significance:

Data science research involves large volumes of data, often derived from unconventional sources. Given the complex nature of big data research, there is a strong need for the development of ethically appropriate protocols that are sensitive to the complexities of data science and data sources. While reviews of health research by research ethics committees are necessary from an ethical and legal perspective, complementary advisory committees such as ethics advisory committees could be established to advise on ethics challenges more broadly. In this Perspective, we describe a multidisciplinary ethics advisory committee linked to a data science research hub in sub-Saharan Africa.

Data science is an interdisciplinary field in which scientific methods, processes, extremely large data sets, machine learning algorithms and information systems are used to extract knowledge and insights from structured and unstructured data.[1,2] The United States National Institutes of Health (NIH)-funded programme, 'Harnessing Data Science for Health Discovery and Innovation in Africa (DS-I Africa)', was established to create a data science research and training network across Africa.[3] The Research for Ethical Data Science in sub-Saharan Africa (REDSSA) project, also NIH-funded, is a unique project planned to complement the focus on data science and its emergence in Africa by exploring the ethical, legal, and social implications (ELSI) of this rapidly growing field. Two of the specific goals of the REDSSA ELSI project are to establish a consortium of sub-Saharan African bioethicists "to develop contextualised guidance in the ELSI of data science" and to establish a Data Science Ethics Advisory Committee (EAC) for the research hubs.[2] The REDSSA project is linked to the INFORM-Africa research hub, operating out of Nigeria. This research hub studies the interaction between SARS-CoV-2 and Human Immunodeficiency Virus (HIV) with the goal of using the data to improve pandemic preparedness.[3]

Ethics review of data science research

Research ethics committees (RECs) traditionally review health research protocols involving human participants with the aims of preventing harm and promoting benefit to research participants, while at the same time ensuring that the research is scientifically valid, and fair, and promotes respect for participants and the community.[4] The focus of the review is on the protection of individuals or groups of participants.

Data science uses big data (large volumes of data), often derived from unconventional sources such as social media, cellular telephone mobility data, wearable technologies, or aggregated health data. Big data have been defined in terms of the three Vs: volume (very large data sets), variety (multiple data formats with structured and unstructured content), and velocity ("high rate of data inflow with non-homogenous structure").[5] Big data in the healthcare context are usually collected for reasons other than research. When accessed for research they are aggregated and deidentified.[6] However, these big data sets are prone to inherent bias as the information is derived from existing data sets. The data analysis yields information relating to types of groups as well as individuals. "The massive scope of big health data coupled with hypothesis-generating interrogation approaches using artificial intelligence (AI) technologies such as machine learning (ML) yields a significant risk of spurious findings."[6] As an example, patients from certain racial groups may be systematically disadvantaged by AI based on cost-of-care data rather than on severity of illness, as the disparity in access to health care skews the algorithm.[7]

The use of big data in research has resulted in a shift of terminology describing the participants as 'data subjects' or 'data sources', rather than the traditional 'human subjects' or 'research participants'. National and international guidelines, legislation, and regulation govern the use and sharing of data to various degrees in different countries in sub-Saharan Africa. In South Africa, legislation such as the *Protection of Personal Information Act* (POPIA)[8] and the *Promotion of Access to Information Act* (PAIA)[9] have also become relevant in research ethics. The concern related to this type of research is that researchers can access potentially sensitive data without any engagement with data subjects. Consequently, the risks relate to 'informational harm' rather than physical harm. Informational harm includes breaches of privacy and 'algorithmic discrimination'.[4,6] The resultant harms are to groups of people as well as individuals, with ensuing emotional distress and discrimination. The complexity of this research means that RECs may not have the expertise to review data science research, or that the research occurs without any involvement of a REC for the review process.[4] An example of the latter was an experiment conducted via Facebook, in which the news feeds of 689 003 users were manipulated to expose them to greater or lesser amounts of emotional content to show that emotional contagion can occur without the awareness of the participants.[10] The study was severely criticised for the lack of ethics oversight and for potentially exposing vulnerable participants to significant harm.[11,12]

Regulatory and governance processes have emerged to establish oversight of new research contexts. As an example, Facebook established an Oversight Board in 2020; however, the value of such oversight has been debatable. Ferretti et al. question whether the growth in oversight mechanisms is simply a tick-box process, "motivated by the urge to fill the existing regulatory gaps, or whether it is just 'ethics washing'"[4]. Oversight mechanisms involved in data sharing, such as data safety monitoring boards and data access committees, are not new and have existed for some time.[4] However, data science review probably requires closer collaboration between RECs and data safety monitoring boards, with RECs becoming more involved in routine monitoring and oversight of data and data access.

Data access committees

Many countries have established guidelines and policies that govern data sharing within and across their borders. Data access committees (DACs) are tasked with protecting the rights and interests of the parties involved in genomic data sharing by reviewing requests for access. There are three types of DACs: (1) DACs in single research groups, where the study's principal investigator (PI) or co-investigator assists with managing requests; (2) DACs in consortia, where the PIs are assisted by legal and/or ethics experts; and (3) DACs attached to institutes, which function centrally and whose members include the necessary experts. The second type of DAC may have a data access officer to manage regular requests, and an advisory committee functioning at a higher level that includes the PIs and legal and ethics experts to advise on more difficult cases or to establish policy. The major reason for data access control is to protect the privacy of data subjects and foster public trust, together with protecting the professional interests of the data creators.[13] A DAC requires a framework for good governance to guide data access decisions.[14]

Solutions to improve ethics review of data science research

Ferretti et al.[4] propose several reforms to improve the ethics review of data science and big data research. These include regulatory reforms such as new guidance for RECs in the form of flow charts on the ethics of such research, procedural reforms with new working and assessment tools, upskilling of REC members in big data knowledge, and inclusion of subject experts as members of the REC or consulting external experts for specific issues.[4] They also suggest the inclusion of other ethics committees, complementary to RECs in the review process, "to assess big data research and provide sectorial accreditation to researchers". The advantages of this would be to lessen the load on RECs and to obtain expert opinions for big data studies. The disadvantages are potential inefficiency of review procedures, erosion of responsibility of the REC, and questions about the role of the REC in big data ethics review.[4]

Establishing an ethics advisory committee for REDSSA

The research strategy for the REDSSA project included early integration of ELSI into data science research conducted at the research hub. Consequently, REDSSA bioethicists attend weekly meetings of the research hub and are immersed in the scientific and ethics challenges arising on the ground. Advice is provided in real time as issues emerge during the conduct of research. In addition, a Data Science Ethics Advisory Committee was established as a more formal structure to inform broader ethics questions in the research hub. This Committee is interdisciplinary, with representation from ethics, law, data science, social science and the community. All the members of this Committee are part of the REDSSA project and are funded through the grant. No additional funding exists for the EAC. It does not function as a REC or an institutional review board. Instead, substantive ethics issues within the hub or its projects can be referred for deliberation. Such referrals can occur before or after submission to a REC or institutional review board to allow deliberation on specific ELSI concerns.[2] As the REDSSA project is linked to the INFORM-Africa research hub in the DS-I Africa Consortium, referrals from the research hub to the EAC do not require additional funding or attract extra charges.

The terms of reference of the EAC were initially written by a small group of REDSSA team members affiliated with the Centre for Medical Ethics and Law at Stellenbosch University, and then refined by discussion at the first meeting of the EAC, held virtually, and subsequently via two rounds of email communication. The functions of the EAC are advisory, consultative and educational, and include development of recommendations and policy review.

The purpose of the EAC is threefold:

1. To promote and uphold respect for the dignity and rights of research participants/data donors/data subjects/data sources of the INFORM-Africa project.
2. To act as a consultative and resource base on data science ethical issues primarily for the research hub and various stakeholders.
3. To help develop ELSI policies and guidelines with relevant consultants as required.

The responsibilities of the EAC are:

1. To fulfil an advisory and consultative role with respect to data science ethics dilemmas in the research environment.
2. To advise on the development of protocols relating to research ethics dilemmas in conjunction with relevant researchers.
3. To clarify concepts around ethics pertaining to surveillance and research using surveillance data.

The interdisciplinary membership of the EAC includes the following: chairs; vice-chairs; data scientists; bioethicists from sub-Saharan Africa; external bioethicists; research ethics committee members; researchers knowledgeable about COVID-19 and HIV research; legal experts; INFORM-Africa Community Advisory Board members; independent members; and the Secretariat. Each position has two appointees (a primary and an alternative). The Chairs are appointed by the PI from the REDSSA management team for an initial period of 1 year, renewable annually thereafter up to a maximum of 3 years. The Chairs are assisted by the Vice-Chairs.

The PI and co-PI of the REDSSA project, together with the NIH Scientific Officer overseeing the REDSSA project, are ultimately responsible for the oversight of the EAC. The members of the EAC are required to declare conflicts of interest and recuse themselves from the discussion as appropriate.

When necessary, ad hoc members – such as relevant bioethics, legal or data science representatives, and experts in specific fields – are consulted on a case-by-case basis.

The REDSSA EAC is an advisory/consultative body and not a decision-making committee. Research requiring ethics approval must be reviewed by an institutional or national REC. The REC could consult experts in the ethics of data science if necessary, including the EAC which would be facilitated by the INFORM-Africa PI.

All committee deliberations remain confidential. The EAC recommendations are formulated via consensus. The recommendations are minuted and forwarded to the INFORM-Africa PI, who may share them with the REC. If the INFORM-Africa PI disagrees with the recommendation of the EAC, the decision and the reason(s) for that decision should be communicated to the EAC in writing. The EAC members discuss how to manage this on a case-by-case basis, but a report is forwarded to the NIH Scientific Officer. The final regulatory approval and oversight of research projects lies with the REC. If the REC disagrees with the advice provided by the EAC, a meeting should be convened between the Chair of the EAC, the Chair of the REC and the PI to discuss the project. Likewise, if a PI seeks advice from the EAC about an approved project, such advice should be communicated to the REC.

A two-part framework for assessing the ethical implications of big data health-related research projects is suggested by investigators from the United Kingdom Research Study into Ethnicity and COVID-19 Outcomes in Healthcare Workers (UK-REACH) study group. Firstly, "the specific legal and ethical issues raised by the project's aims and methods" must be identified; and, secondly, those issues must be addressed to promote positive aspects of the project (such as justice and respect for persons) while simultaneously modifying or eradicating negative aspects (such as stigmatisation, legal violations, and aggravation of social inequality and injustice).[15] A framework for assessing the ethics of big data research is described by Xafis and co-authors[16] in which they detail the following steps: (1) Identify and explicate the moral and ethical issue(s); (2) Identify the relevant values (both substantive and procedural) pertaining to the issue or problem; (3) Identify potential solutions and actions; (4) Evaluate the relative ethical importance of the different options; (5) Select the option that carries the greatest ethical weight while considering the roles and influence of the decision-makers in the group; and (6) Communicate

the decision clearly to all stakeholders. The REDSSA EAC deliberates on potential ethics frameworks that best suit its decision-making needs.

Conclusion

Data science has the potential to enhance health-related knowledge, particularly in the field of public health. However, big data projects are subject to ethical and legal concerns and RECs may experience challenges with the review process as data ethics is an emergent discipline in sub-Saharan Africa. EACs may play a supportive role in big data research for both researchers and RECs. The REDSSA EAC provides one viable way of fulfilling an advisory role that can better support researchers and RECs involved in big data research.

Competing interests

There are no competing interests to declare. The authors are all members of the REDSSA EAC.

References

1. Dhar V. Data science and prediction. Communications of the ACM. 2013;56(12):64–73. https://doi.org/:10.1145/2500499
2. Moodley K, Rennie S. Research strategy of REDSSA. In: Unpublished NIH grant application.
3. Munyaradzi M. New funding set to transform data science research and innovation. Nat Afr. 15 November 2021 [cited 2022 Sep 02]. https://doi.org/10.1038/d44148-021-00111-3
4. Ferretti A, Ienca M, Sheehan M, Blasimme A, Dove ES, Farsides B, et al. Ethics review of big data research: What should stay and what should be reformed? BMC Med Ethics. 2021;22, Art. #51. https://doi.org/10.1186/s12910-021-00616-4
5. Sivarajah U, Kamal MM, Irani Z, Weerakkody V. Critical analysis of big data challenges and analytical methods. J Bus Res. 2017;70:263–286. https://doi.org/10.1016/j.jbusres.2016.08.001
6. Doerr M, Meeder S. Big health data research and group harm: The scope of IRB review. Ethics Hum Res. 2022;44(4):34–38. https://doi.org/10.1002/eahr.500130
7. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366:447–453. https://doi.org/10.1126/science.aax2342
8. Protection of Personal Information Act (POPIA) [homepage on the Internet]. c2019 [cited 2022 Sep 02]. Available from: https://popia.co.za/
9. Promotion of Access to Information Act 2 of 2000 (PAIA) [webpage on the Internet]. c2000 [cited 2022 Sep 02]. Available from: https://www.gov.za/documents/promotion-access-information-act
10. Adam D, Kramer I, Guillory JE, Hancock JT. Experimental evidence of massive scale emotional contagion through social networks. Proc Natl Acad Sci USA. 2014;111(24):8788–8790. https://doi.org/10.1073/pnas.1320040111
11. Vitak J, Shilton K, Ashktorab Z. Beyond the Belmont principles: Ethical challenges, practices, and beliefs in the online data research community. Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing; 2016 February 27 – March 02; San Francisco, CA, USA. New York: Association for Computing Machinery; 2016. p. 941–953. http://dx.doi.org/10.1145/2818048.2820078
12. Facebook faces UK probe over emotion study. BBC World. 2014 July 02 [cited 2022 Sep 02]. Available from: https://www.bbc.com/news/technology-28102550
13. Shabani M, Borry P. "You want the right amount of oversight": interviews with data access committee members and experts on genomic data access. Genet Med. 2016;18(9):892–897. https://doi.org/10.1038/gim.2015.189
14. De Vries J, Abayomi A, Littler K, Madden E, McCurdy S, Ouwe Missi Oukem-Boyer O, et al. Addressing ethical issues in H3Africa research – the views of research ethics committee members. The HUGO J. 2015;9, Art. #1. https://doi.org/10.1186/s11568-015-0006-6)
15. Reed-Berendt R, Dove ES, Pareek M. The ethical implications of big data research in public health: "Big Data Ethics by Design" in the UK-REACH Study. Ethics Hum Res. 2022;44(1):2–17. https://doi.org/10.1002/eahr.500111
16. Xafis V, Schaefer GO, Labude MK, Brassington I, Ballantyne A, Lim HY, et al. An ethics framework for big data in health and research. Asian Bioeth Rev. 2019;11:227–254. https://doi.org/10.1007/s41649-019-00099-x