

**AUTHOR:**

Gonasagrie Nair¹ 
 Theresa L. Burgess^{1,2} 
 Adetayo E. Obasa¹ 
 Sharon Kling^{1,3} 
 Shenuka Singh^{1,4} 

AFFILIATIONS:

¹Centre for Medical Ethics and Law, WHO Bioethics Collaborating Centre, Department of Medicine, Stellenbosch University, Cape Town, South Africa

²Division of Physiotherapy, Department of Health and Rehabilitation Sciences, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

³Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

⁴Discipline of Dentistry, School of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

CORRESPONDENCE TO:

Gonasagrie Nair

EMAIL:

lulu.nair13@gmail.com

HOW TO CITE:

Nair G, Burgess TL, Obasa AE, Kling S, Singh S. Revisiting community engagement methods in the context of data science research and big data use in South Africa. *S Afr J Sci.* 2023;119(5/6), Art. #14723. <https://doi.org/10.17159/sajs.2023/14723>

ARTICLE INCLUDES:

- Peer review
- Supplementary material

KEYWORDS:

data science, big data, community engagement, data ownership, citizen science

PUBLISHED:

30 May 2023

FUNDING:

US National Institutes of Health (U01MH127704)



Revisiting community engagement methods in the context of data science research and big data use in South Africa

Significance:

Effective community engagement for the use of large data sets in health research is faced with challenges similar to those in investigator-driven research. The scope of community engagement has evolved in high-income countries to embrace citizen science by communities and regulators to build trust in data science research. In South Africa and other low- and middle-income countries, with varying levels of literacy and the influence of pre-existing beliefs and past negative experiences with research, advisory committees of diverse stakeholder composition still have a role to play in protecting the rights of researched communities.

Introduction

Access to existing large diverse data sets plays an important role in drug development research, precision medicine, diagnostic imaging, artificial intelligence (AI) platforms, medical decision support systems, and managing public health emergencies.¹

Large volumes of genomic and phenotypic health-related data are collected from various sources including computers, smartphones, tablets, and wearable devices. Electronic health data are also collected by medical insurance companies in the private health sector and from public health data bases.² These data are categorised as ‘big data’, given that the information originates from a variety of sources, is of large volume, and is processed at high speeds.³ Data science, which makes use of big data, is defined as the ‘study of the extraction of knowledge from data’ and differs from statistics because the data sources and formats vary.⁴ Data may be presented as numbers, text, images, or video. A multidisciplinary approach, involving computer scientists, sociologists, clinicians and epidemiologists, is required for the analysis and interpretation of health data.⁴

Importantly, access to large pre-existing data sets may increase the efficiency of research by avoiding potential duplication and overburdening research participants and increasing statistical power and the generalisability of study findings.

While the value of data science cannot be underscored, there may be a lack of awareness among the public⁵, especially in low- and middle-income countries (LMICs), that these data are being collected and shared with in-country researchers or with researchers in other countries. At the same time, there are ethical and legal challenges associated with health data science research that need to be considered, particularly to maintain individual patient and community trust in research. This emphasises the need for sustained community and stakeholder engagement by researchers. In this Perspective, we therefore highlight the ethical implications of big data research, the use of community and stakeholder engagement to build data science literacy and public trust, the limitations of traditional community engagement, especially in LMICs and South Africa, and how these identified challenges could be addressed.

Ethical considerations

The use of routine clinical data for research purposes results in a blurring of boundaries between clinical care and research, and raises questions around data ownership, patient privacy, and autonomy.⁶

Some of the possible harms to research participants could include violation of privacy, and stigma based on health-seeking behaviour and health patterns of communities. Additionally there could be secondary discrimination from data sets used to generate algorithms, which could lack diversity and thereby introduce bias in the interpretation of the study findings.⁷ Apart from issues of privacy and confidentiality, questions of data ownership arise if health-related data have already been collected as part of routine clinical care and have subsequently been shared for research purposes.

From a consent perspective, while clinical research allows for broad consent or tiered consent for the storage of samples and use of data for future related research, consent for clinical care is typically only for specific clinical management. Additionally, data could be accessed purely for clinical purposes and later re-purposed for research, yet consent was only obtained for the clinical services provided.

Legislation related to consent for data use in South Africa adds to this unclear picture for big data research. As per the *South African National Health Act*⁸ and Health Professionals Council of South Africa guidelines⁹, a patient has the right to expect that their health-related data will be confidential and that sharing of this information will only occur after their consent has been obtained. In contrast, the *Protection of Personal Information Act of 2013*¹⁰ allows for the sharing of special information, that is, health data, if these are de-identified.

The need for community engagement

The highlighted ethico-legal considerations reiterate the need for better engagement with affected individuals and communities. Community engagement is seen as a vital process to optimise public trust in the research process.^{11,12} The levels of community engagement include stakeholder input, consultation, collaboration, and shared leadership, with shared leadership being the most collaborative and stakeholder input being least so.¹³



However, community engagement may also result in unintended consequences. Although researchers may have good intentions to initiate meaningful conversations with research communities around the research and protocol development, such efforts may be misinterpreted and intentions may be misconstrued, thereby leading to mistrust between the researched communities and the research teams involved.¹⁴

There are several reasons related to non-participation in research, including a lack of understanding of the research, and considering the research irrelevant either because of a feeling that it does not address the needs of communities or by misinterpreting it as being elitist. Past negative experiences and/or cultural barriers may also play a role.⁵ Community engagement has assisted in identifying these reasons and addressing them in a culturally sensitive manner to allow for research participation, and has thus been beneficial both to the researcher in allowing successful trial implementation and to the community in addressing health priorities. The success of research is dependent not only on the occurrence of community engagement, but as communities become more familiar with the advantages and pitfalls of participating in research, by the extent of community engagement.⁵

Limitations in the current frameworks to guide community engagement in data science

Frameworks for participatory research have been developed to promote authentic community engagement through a sense of ownership and to meet funder/s' requirements.¹¹ These frameworks, although useful, are not formally recognised by policymakers or research ethics committees (RECs). In settings in which community engagement is not mandatory or required, original participatory engagement intentions fall away and, due to mistrust and disillusionment, communities with capacity shift from collaborative engagement to one of capacitation, where the community identifies research priorities, implements research, analyses data and disseminates results.⁵ This shift to the 'capacitation' model, which is being recommended and adopted in high-income countries^{5,6,13}, may not be feasible in LMICs due to the scarcity of human and financial resources for sustainability.⁵

Additionally, strategic plans for health research highlight the value of community engagement⁵, but there are no identified processes to enforce its implementation. The Emanuel, Wendler, and Grady framework has been adopted by some RECs globally for evaluation of ethical, social, and scientific robustness of proposed research and a 2008 revision included collaborative partnership for the first time.¹⁶ In spite of this recommendation that collaborative partnership is one of the eight factors considered in ethics review, a South African study indicated collaborative partnership was less likely to raise queries during the review process in comparison to the other factors, if considered at all.¹⁷

In comparison to health research focused on a specific disease or condition, the community in health data science research is not as clearly defined. If we consider the diverse sources of big data, questions around who constitutes the community and stakeholders arise. All users of social media, owners of a cell phone or wearable device and those who seek health care in either the private or public sectors may be considered the research community. However, narrowing the health-seeking behaviour to a particular health condition such as Human Immunodeficiency Virus (HIV), tuberculosis (TB), or to a rare disease will facilitate the identification of the community even in big data research.⁵

Traditional mechanisms for community engagement within the context of health research and clinical trials have involved community advisory boards (CABs). CABs generally constitute influential community members who serve as a bridge between community members and researchers, thus ensuring optimal study implementation and protection of the rights of communities.

The need for a paradigm shift

Funders and governance structures of clinical research in high-income countries with stringent data protection and protection of personal information laws are requiring comprehensive involvement of patients, as owners of their health data, in research.^{6,12} This has seen the advent

of greater degrees of citizen science, with patients deciding for which projects their data will be used and uploading data directly to databases.⁶

Our proposal

Ferretti et al.¹⁸ state that the use of big data excludes engagement with study participants, but we are of the opinion that the CAB model for community engagement would still be appropriate in addition to models that encompass more participatory methods of community engagement. One such participatory approach would be 'crowdsourcing' – characterised by large groups of experts and non-experts from diverse backgrounds providing solutions to a problem. This is an approach that can be used in clinical health research.¹⁹

Where there are well-defined accessible communities, the ethical principles that govern research can be adhered to through CABs and REC review. CAB review of consent forms to ensure social and cultural appropriateness and advice on the consent process ensures respect for the autonomy of study participants. However, data science research involves the re-use of pre-existing data sets so consent is not sought from individuals or communities, but new mechanisms, such as dynamic or portable consent made possible through online platforms, may be a solution.²⁰ Protocol review, prior to study implementation, ensures that principles of beneficence/non-maleficence and justice are adhered to. This ensures that ineligible participants are linked to care, that benefits outweigh risks, that study participants are not required to waive any of their rights, and that post-trial access and benefit sharing mechanisms are in place to ensure access to successful interventions to those who endured the risks of study participation. While this approach may be regarded as paternalistic and non-empowering, it still has a role in research-naïve communities and many indigent communities in LMICs in which individuals may be coerced into study participation. Ferretti et al.⁷ note that RECs, which often include a community representative, may struggle to apply existing governance frameworks or regulatory tools for ethics review for data science research because data are anonymised and the research does not involve interaction with research participants. We agree with these authors. There is a need to build further capacity in RECs with regard to the review of protocols related to big data science. Such capacity building should include ongoing educational training as well as ensuring that RECs include members with appropriate skills and experience in this evolving area of research.

Conclusion

Community engagement for health research utilising large data sets should include public engagement or 'data science citizenship'. However, there is a role for traditional engagement to foster trust and transparency through CABs where stakeholders are existing research participants. RECs should be empowered to critically evaluate community engagement in data science health research. In-country regulations for data ownership and sharing should align with each other for easy interpretation by communities and researchers, both local and international.

Competing interests

We have no competing interests to declare.

References

1. Mehta N, Shukla S. Pandemic analytics: How countries are leveraging big data analytics and artificial intelligence to fight COVID-19? *SN Comput Sci.* 2022;3, Art. #54. <https://doi.org/10.1007/s42979-021-00923-y>
2. Vayena E, Blasimme A. Biomedical big data: New models of control over access, use and governance. *J Bioeth Inq.* 2017;14:501–513. <https://doi.org/10.1007/s11673-017-9809-6>
3. Benedek A, Molnar G, Szuts Z. Practices of crowdsourcing in relation to big data analysis and education methods. In: 2015 IEEE 13th International Symposium on Intelligent Systems and Informatics (SISY); 2015 September 17–19; Subotica, Serbia. IEEE; 2015. p. 167–172. <https://doi.org/10.1109/SISY.2015.7325373>
4. Dhar V. Data science and prediction. *Commun ACM.* 2013;56:64–73. <https://doi.org/10.1145/2500499>



5. Grayson S, Doerr M, Yu JH. Developing pathways for community-led research with big data: A content analysis of stakeholder interviews. *Heal Res Policy Syst.* 2020;18, Art. #76. <https://doi.org/10.1186/s12961-020-00589-7>
6. Winickoff DE, Jamal L, Anderson NR. New modes of engagement for big data research. *J Responsible Innov.* 2016;3:169–177. <https://doi.org/10.1080/23299460.2016.1190443>
7. Ferretti A, Ienca M, Velarde MR, Hurst S, Vayena E. The challenges of big data for research ethics committees: A qualitative Swiss study. *J Empir Res Hum Res Ethics.* 2022;17:129–143. <https://doi.org/10.1177/15562646211053538>
8. National Health Act 61 of 2003 (NHA) [webpage on the Internet]. c2003 [cited 2022 Sep 05]. Available from: <https://www.gov.za/documents/national-health-act>
9. Health Professions Council of South Africa (HPCSA). General ethical guidelines for good practice in telehealth: Booklet 10 [document on the Internet]. c2021 [cited 2022 Sep 05]. Available from: https://www.hpcsablogs.co.za/wp-content/uploads/2022/08/Booklet-10_Telehealth_Dec_2021_.pdf
10. Protection of Personal Information Act (POPI Act). Available from: <https://popia.co.za/>
11. Moodley K, Beyer C. Tygerberg Research Ubuntu-Inspired Community Engagement Model: Integrating community engagement into genomic biobanking. *Biopreserv Biobank.* 2019;17:613–624. <https://doi.org/10.1089/bio.2018.0136>
12. US National Institute for Health Research. Going the extra mile: Improving the nation's health and wellbeing through public involvement in research [document on the Internet]. c2015 [cited 2022 Sep 05]. Available from: https://www.researchgate.net/publication/279180522_%27Going_the_extra_mile_improving_the_nation%27s_health_and_wellbeing_through_public_involvement_in_research%27
13. Patient-Centered Outcomes Research Institute (PCORI). The value of engagement [webpage on the Internet]. c2018 [cited 2022 Sep 06]. Available from: <https://www.pcori.org/engagement/value-engagement>
14. Wilkinson A, Slack C, Thabette S, Salzwedel J. "It's almost as if stakeholder engagement is the annoying 'have-to-do'...": Can ethics review help address the "3 Ts" of tokenism, toxicity, and tailoring in stakeholder engagement? *J Empir Res Hum Res Ethics.* 2022;17:292–303. <https://doi.org/10.1177/15562646221078415>
15. South African Department of Health. National Health Research Strategy: Research priorities for SA – 2021–2024. No date [cited 2022 Aug 26]. Available from: <https://www.health.gov.za/wp-content/uploads/2022/05/National-Health-Research-Priorities-2021-2024.pdf>
16. Emanuel EJ, Wendler DD, Grady C. In: Emanuel EJ, Grady C, Crouch RA, Lie RK, Miller FG, Wendler DD, editors. *The Oxford textbook of clinical research ethics.* Oxford/New York: Oxford University Press; 2008. p. 123–135.
17. Tsoka-Gwegweni JM, Wassenaar DR. Using the Emanuel et al. framework to assess ethical issues raised by a biomedical research ethics committee in South Africa. *J Empir Res Hum Res Ethics.* 2014;9:36–45. <https://doi.org/10.1177/1556264614553172>
18. 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>
19. Tucker JD, Day S, Tang W, Bayus B. Crowdsourcing in medical research: Concepts and applications. *PeerJ.* 2019;7, e6762. <https://doi.org/10.7717/peerj.6762>
20. Cato KD, Bockting W, Larson E. Did I tell you that? Ethical issues related to using computational methods to discover non-disclosed patient characteristics. *J Empir Res Hum Res Ethics.* 2016;11:214–219. <https://doi.org/10.1177/1556264616661611>