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 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 databases.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 public, 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 Act5 and Health Professionals Council of South Africa guidelines,8 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 20139 allows for the sharing of special information, that is, health data, if these are de-identified.

The need for community engagement
The highlighted ethical 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

© 2023. The Author(s). Published under a Creative Commons Attribution Licence.
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 funders’ 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 capitulation, where the community identifies research priorities, implements research, analyses data and disseminates results.2 This shift to the ‘capacitation’ model, which is being recommended and adopted in high-income countries5,13, 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 engagement14, 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.11 In spite of this recommendation that collaborative partnership is one of the eight factors considered in ethics review, a South African study indicated that collaborative partnership is one of the factors 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


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

