

**AUTHORS:**

Nezerith Cengiz¹ 
 Siti M. Kabanda¹ 
 Tonya M. Esterhuizen² 
 Keymanthri Moodley¹ 

AFFILIATIONS:

¹Centre for Medical Ethics and Law, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

²Division of Epidemiology and Biostatistics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

CORRESPONDENCE TO:

Nezerith Cengiz

EMAIL:

ncengiz@sun.ac.za

DATES:

Received: 30 Sep. 2022

Revised: 01 Feb. 2023

Accepted: 02 Mar. 2023

Published: 30 May 2023

HOW TO CITE:

Cengiz N, Kabanda SM, Esterhuizen TM, Moodley K. Exploring perspectives of research ethics committee members on the governance of big data in sub-Saharan Africa. *S Afr J Sci.* 2023;119(5/6), Art. #14905. <https://doi.org/10.17159/sajs.2023/14905>

ARTICLE INCLUDES:

- Peer review
- Supplementary material

DATA AVAILABILITY:

- Open data set
- All data included
- On request from author(s)
- Not available
- Not applicable

EDITOR:

Pascal Bessong 

KEYWORDS:

big data, data governance, data regulation, research ethics committees, sub-Saharan Africa

FUNDING:

US National Institutes of Health (1U01MH127704-01)



Exploring perspectives of research ethics committee members on the governance of big data in sub-Saharan Africa

Interest in the governance of big data is growing exponentially. However, finding the right balance between making large volumes of data accessible, and safeguarding privacy, preventing data misuse, determining authorship and protecting intellectual property remain challenging. In sub-Saharan Africa (SSA), research ethics committees (RECs) play an important role in reviewing data-intense research protocols. However, this regulatory role must be embedded in a context of robust governance. There is currently a paucity of published literature on how big data are regulated in SSA and if the capacity to review protocols is sufficient. The aim of this study was to provide a broad overview of REC members' awareness and perceptions of big data governance in SSA. A descriptive cross-sectional survey was conducted from April to July 2022. We invited 300 REC members to participate in our online survey via Research Electronic Data Capture (REDCap). A total of 140 REC members, representing 34 SSA countries, completed the online survey. Awareness of data governance laws, policies and guidelines was variable across the subcontinent. A quarter of respondents (25%) indicated that national regulations on the trans-border flow of research data are inadequate. Institutional policies on research data protection were also regarded as being inadequate. Most respondents (64%) believed that they lacked experience in reviewing data-intense protocols. Data governance and regulation in SSA need to be strengthened at both national and institutional levels. There is a strong need for capacity development in the review of data-intense research protocols on the subcontinent.

Significance:

This is the first empirical survey in SSA in which awareness and perspectives of REC members have been explored specifically relating to the review of data-intense research protocols. Big data have raised new ethics and legal challenges, and this survey provides a broad overview of these challenges in SSA. Our study confirms that knowledge and awareness of legislative frameworks and ethics guidance in SSA vary considerably where big data are concerned. The research results could be useful for a range of stakeholders, including RECs, data scientists, researchers, research and academic institutions, government decision-makers and artificial intelligence (AI) coders.

Background

The abundance of health and research data that exists today has enormous potential to unlock future advances in science – a prospect discussed for decades by researchers and policymakers.¹ Recently, the potential of big data to solve some of the world's most challenging problems has become more apparent. 'Big data' refers to large volumes of a variety of raw data processed at high speed and frequency.² The sharing of research data is of increasing interest, with many funders advocating for, or even requiring researchers to share data sets as a condition of funding to maximise their utility and value.³ Understandably, sharing research data is regarded as a best practice by the World Health Organization (WHO).^{4,5}

Despite the benefits of data sharing, finding the right balance between making data accessible and safeguarding privacy, preventing data misuse, determining authorship and protecting intellectual property is challenging.^{4,6,7} This challenge has been reported to be greater in low- and middle-income countries (LMICs) such as in sub-Saharan Africa (SSA) because of the gap that exists in decision-making between data producers and data users.^{4,7} Some SSA countries have introduced data protection regulations in response to the recent digital revolution.

South Africa is one of the countries that has sought to enforce data governance via the *Protection of Personal Information Act (POPIA), Act No. 4 of 2013*, which came into force on 1 July 2020.⁸ However, legal and ethics frameworks to guide data sharing and protect the interests of data donors on the subcontinent appear to vary considerably in their structure, terms, procedures and authority.⁹

Data protection has also become concerning in the context of the cross-border transfer of human biological materials (HBMs) and data.¹⁰ In response to this, Material Transfer Agreements (MTAs) and Data Transfer Agreements (DTAs) have evolved to contractually govern the transfer of biological materials and data between parties to protect the interests of stakeholders.¹¹ A DTA is a legal contract governing the transfer of deidentified human subject data, or identifiable human subject data in cases where a respondent has given voluntary, informed and electronic consent.¹² DTAs are required when data owned by one institution are transferred to another institution for the continuation of research efforts. A DTA sets out the related protection, rights and obligations of both parties and delineates the specific purpose(s) for which the data may be used. This facilitates the cross-border transfer of data.^{11,12} In some countries, there is an additional requirement to inform the relevant national data protection authority about the cross-border transfer of data.

Research ethics committees (RECs) have traditionally been established to protect the rights of research participants. However, they also play an important role in reviewing data-intensive research protocols where data protection and data sharing are important.¹³ The recent pandemic has placed increasing demands on RECs as research engaging with big data and artificial intelligence (AI) was accelerated. Many scholars have been deliberating on the role of RECs in reviewing data-intensive research protocols, and have found that developed countries such as Switzerland², the UK¹⁴ and Australia¹⁵ lack the expertise or skills to review such studies. Big data research should be differently legislated and considered as it poses greater or unique risks and implications than flows of samples. Conventional informed consent is not ideal for protecting participants in big data research.² Other examples of the implications of big data research include anonymisation, algorithmic bias, data protection, data storage and data reuse. In many countries in SSA, biological samples are regulated in legislation via MTAs and in guidelines.¹⁶ However, data, and particularly big data, are excluded. The rapid flow of large volumes of data carries benefits to science, but also many risks to personal information protection and governance, and should be regulated.

The data ecosystem is becoming increasingly complex. Apart from RECs, Data Access Committees (DACs) have emerged as another governance mechanism to manage the controlled access of data.¹³ A DAC comprises a group of individuals who have the responsibility of reviewing and assessing research data access requests.¹³ They may serve as part of an REC or may be an independent committee in an institution or country with the aim of promoting the benefits of data access, whilst minimising potential harm to data respondents or donors.¹³

Data governance is understood as the practice of safeguarding valuable information from exploitation, compromise and loss or theft. It is largely executed through regulatory and legal data protection frameworks.¹⁷⁻¹⁹ These frameworks govern how certain data types are collected, processed and shared. This secures the privacy, availability and integrity of data through frameworks that set out how sensitive data, in particular, and privacy should be managed via the provision of tools and policies that restrict the unauthorised access, use and/or transfer of data.¹⁷⁻¹⁹ Examples of personal identifiable data include names, photographs, email addresses, bank account details, the Internet Protocol (IP) addresses of personal computers and biometric data.¹⁷

It is important to note that data protection laws may differ across various countries, thereby causing an inequality and disparity in the degree of data protection. Some of these countries have stricter rules that apply, which may require notification or approval by the data protection authority and/or special conditions, as well as consent from the data subject as a requirement for the cross-border transfer of data.²⁰

In South Africa, the National Health Research Ethics Council (NHREC) developed a national guideline, 'Ethics in Health Research: Principles, Processes and Structures', in 2015 to ensure that research is conducted responsibly and ethically in South Africa.²¹ The NHREC emphasises the importance of recognising the values, beliefs and attitudes of data donors.²¹

The guidance document recommends the responsible management of data collection, informed consent, the protection of vulnerable populations, the permissible secondary use of data, and the non-maleficent use of genetic and genomic research.²¹ However, these guidelines are not specific to big data collection, and improved recommendations are required to meet international standards of data management.^{21,22}

Being cognisant of the challenges in the big data ecosystem in LMICs, we aimed to determine REC members' perceptions of data governance in SSA and to describe related challenges. This study is part of a bigger project exploring the ethical, legal and social implications of big data and AI in SSA.

To date, there are no published studies from SSA that have explored the perspectives of REC members on data governance or on the review of data-intensive research protocols. Consequently, it is unclear how REC members on the subcontinent navigate governance structures and processes, and review such protocols. This study offers a novel contribution to the empirical literature in SSA as it aimed to explore these perspectives.

Methods

Study design and sampling

A descriptive cross-sectional survey with both quantitative and qualitative components, involving 140 REC members representing 34 SSA countries, was conducted from April to July 2022. Our aim was to recruit at least one representative from each of the 49 SSA countries. The study population was selected based on membership of a private, institutional or national REC in SSA.

Respondents were invited to participate in an online survey through a web-based application, Research Electronic Data Capture (REDCap). We recruited our sample of REC members through a purposive selection of professional networks of the Stellenbosch University's Centre for Medical Ethics and Law across SSA, and employed a snowballing technique to recruit further respondents.²³ All respondents participated in their personal capacities and provided online consent prior to their completion of the survey.

The survey instruments

The questionnaire was developed based on a review of the literature and consultation with experts in research ethics. A final draft of the questionnaire was developed using REDCap. This online questionnaire was piloted with six REC members from Stellenbosch University to assess its legibility, eliminate ambiguous questions, address repetition and identify any missing information. This was to ensure the face validity of the data collection tool.

The piloted version of the questionnaire consisted of 20 closed-ended questions, of which four were conditional questions that required respondents to meet a certain condition to be asked the following question. These questions were used to establish baseline data regarding the existence of research data-sharing frameworks and guidelines in SSA, the level of awareness of these frameworks and guidelines by REC members, and perspectives regarding existing legal and ethical challenges. In the questionnaire, we distinguished between the institutional and national governance of research data protection and the trans-border flow of research data to take into account the SSA countries without national governance laws. These were divergent across some institutions and countries.

The data collection tool was developed in English and translated into French and Portuguese to cater for Francophone and Lusophone countries.

Data analysis

Survey responses were exported from REDCap into the Statistical Package for Social Sciences (SPSS) version 28 for analysis. Frequencies and percentages were used to describe responses to the closed questions. A trained researcher analysed the answers from the open-ended questions manually by identifying recurring responses.

Ethical aspects

Research integrity was maintained throughout the study, and participation in this research remained entirely voluntary. This survey was a minimal-risk study as the questionnaires involved a factual enquiry with educated, empowered respondents who had the full capacity to consent or decline participation. We approached members in their individual capacities, and respondents consented in their personal capacities. Ethics approval was granted by the Health Research Ethics Committee of the Faculty of Medicine and Health Sciences (reference no: N22/03/028) at Stellenbosch University, South Africa.

Results

Demographic information

A total of 300 individuals were invited to participate in the research study and 140 completed the online survey, yielding an overall response rate of 47% (140/300). The total number of respondents represented 34 of the 49 SSA countries (Figure 1).

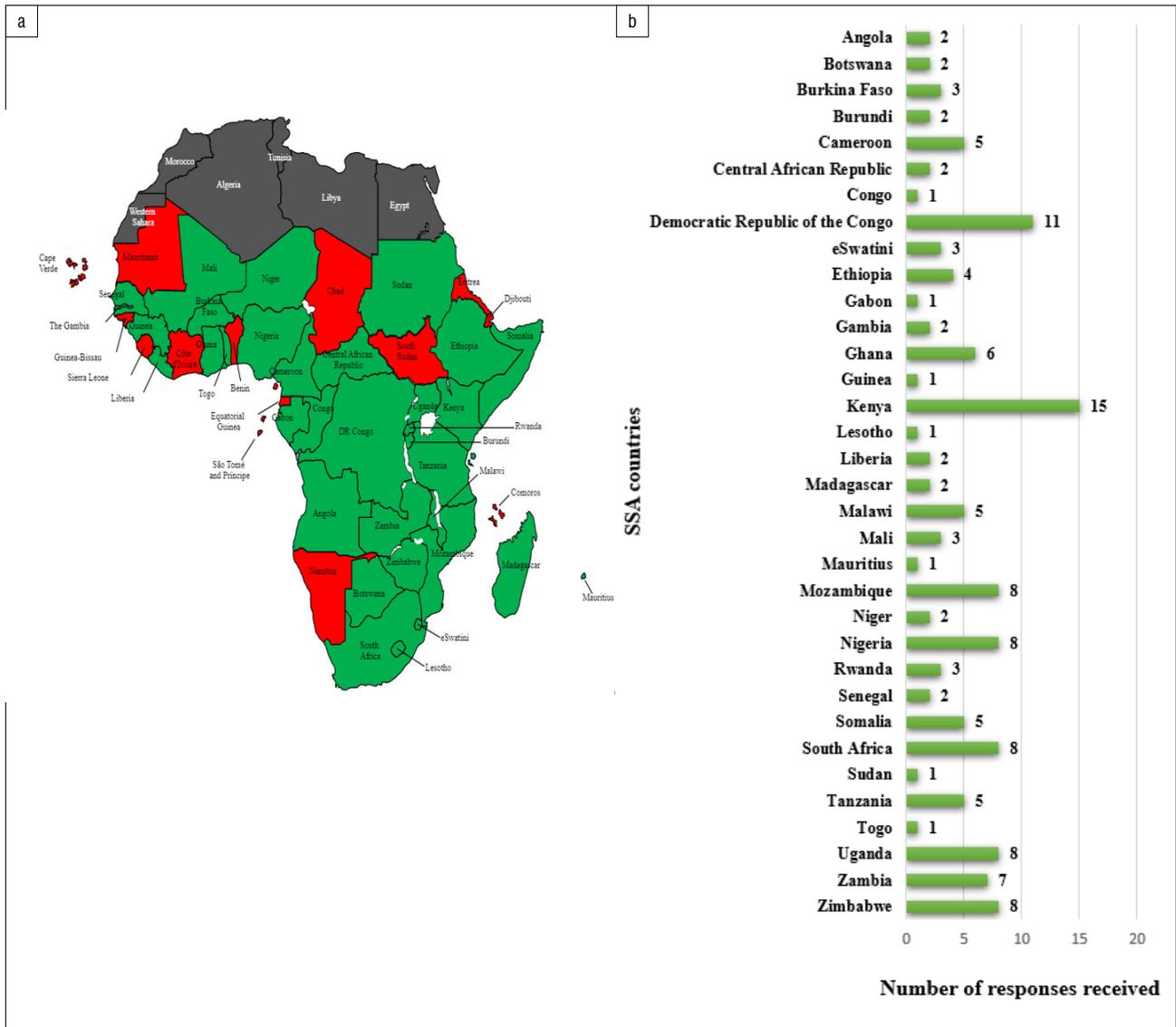


Figure 1: (a) Sub-Saharan Africa (SSA) and (b) the representation of responses received across SSA countries.

More than half the respondents (63%) self-identified as male (88/140), whilst 46% of the respondents (64/140) were PhD graduates, and 41% (58/140) were master's degree graduates (Table 1). Of the respondents, 80% (112/140) had served in the capacity of an REC member and most responses (69%) came from those who had served on an institutional REC (96/140).

Awareness of current laws and policies on research data protection

Just over half the respondents (59%; 82/140) indicated that their country had laws on research data protection (Table 2). Less than half (48%; 67/140) indicated that their country had restrictions and/or prohibitions regarding the trans-border flow of research data. We validated whether respondents responded correctly when reporting on the existence of legislation in their respective countries (Table 3). Of 107 respondents, 76% (81/107) showed concordance, whilst 24% (26/107) showed discordance. For this calculation, we excluded the 33 'unsure' responses. The validity, estimated at 76% in the study, was based on this one question.

Most respondents (69%; 96/140) indicated that their institutions had policies on research data protection, and 50% (70/140) specified that restrictions and/or prohibitions for the trans-border flow of research

data were also in place. Interestingly, just over a third (34%) of the respondents (48/140) mentioned that their affiliated institutions had no restrictions for the trans-border flow of research data.

Perceptions of the current laws and policies on research data protection and transfer

Respondents were asked to indicate how much they agreed or disagreed (on a six-point scale) with statements about the adequacy of their country's laws and institutional policies on research data protection (Table 4). Of the respondents, 45% (63/140) expressed the view that their country's current laws on research data protection were adequate, whereas 19% (27/140) disagreed. Of those who disagreed, 9% (12/140) disagreed strongly. Similarly, 40% (56/140) of respondents perceived their national restrictions and prohibitions on the trans-border flow of research data to be adequate. Of those who agreed, only 7% (10/140) agreed strongly. Just over half (51%) of all respondents (72/140) perceived their institutional policies on research data protection to be adequate.

On the other hand, a quarter (25%) of the respondents (35/140) indicated that their national restrictions and prohibitions on the trans-border flow of research data were inadequate. Slightly fewer (21%; 29/140) felt that their institutional policies on research data protection were also inadequate.

Table 1: Characteristics of survey respondents ($N = 140$)

Characteristic	<i>n</i> (%)
Gender	
Male	88 (63)
Female	51 (36)
Other	1 (1)
Education/qualification	
Bachelor's degree	7 (5)
Honours degree	6 (4)
Master's degree	58 (41)
Doctoral degree	64 (46)
Other	5 (4)
Type of REC	
National	37 (26)
Institutional	96 (67)
Private	7 (5)
Position/role	
Chair	19 (14)
Co-chair	4 (3)
Vice-chair	5 (4)
Member	112 (80)
Years of experience	
Less than 2 years	20 (14)
2–4 years	19 (14)
4–6 years	29 (21)
6–8 years	13 (9)
8 or more years	59 (42)

Table 2: Respondents' awareness of laws and policies on research data protection ($N = 140$)

Any law on research data protection in the country?	<i>n</i> (%)
Yes	82 (59)
No	24 (17)
Unsure	34 (24)
Restrictions/prohibitions placed on the trans-border flow of research data in the country?	
Yes	67 (48)
No	34 (24)
Unsure	39 (28)
Any policy on research data protection at the institution?	
Yes	96 (69)
No	26 (19)
Unsure	18 (13)
Restrictions/prohibitions placed on the trans-border flow of research data at the institution?	
Yes	70 (50)
No	48 (34)
Unsure	22 (16)

Table 3: Validation of responses received ($N = 140$)

Responses	Existing privacy laws		Total
	Yes	No	
Yes	67/107 (63%)	16/107 (15%)	83/107 (78%)
No	10/107 (9%)	14/107 (13%)	24/107 (22%)
Total	77/107 (72%)	30/107 (28%)	107 (100%)

Table 4: Respondents' perceptions of data-related laws or policies ($N = 140$)

Adequate law on research data protection within the respondent's country	<i>n</i> (%)
None (no law or policy)	25 (18)
Disagree strongly	12 (9)
Disagree somewhat	15 (11)
Unsure	25 (18)
Agree somewhat	46 (33)
Agree strongly	17 (12)
Adequate restrictions or prohibitions on the trans-border flow of research data at country level	
None (no law or policy)	16 (11)
Disagree strongly	17 (12)
Disagree somewhat	18 (13)
Unsure	33 (24)
Agree somewhat	46 (33)
Agree strongly	10 (7)
Adequate institutional-level policy on research data protection	
None (no law or policy)	17 (12)
Disagree strongly	8 (6)
Disagree somewhat	21 (15)
Unsure	22 (16)
Agree somewhat	54 (39)
Agree strongly	18 (13)
Adequate institutional-level restrictions or prohibitions on the trans-border flow of research data	
None (no law or policy)	22 (16)
Disagree strongly	12 (9)
Disagree somewhat	18 (13)
Unsure	31 (22)
Agree somewhat	46 (33)
Agree strongly	11 (8)

Transfer agreements

Awareness of MTAs and DTAs was generally good, but around 20% of respondents (28/140) were uncertain of the existence of such agreements. Just over a third (36%; 50/140) indicated that their institutions had a

separate DTA in place. Most respondents (74%; 103/140) indicated that their REC was required to review DTAs and MTAs. Only 13% (18/140) indicated that their REC did not review these documents (Table 5).

Table 5: Protection of research data or HBM (N = 140)

Separate DTA available at the respondent's institution	n (%)
Yes	50 (36)
No	90 (64)
Separate MTA available at the respondent's institution	
Yes	74 (53)
No	66 (47)
Combined DTA and MTA available at the respondent's institution	
Yes	33 (24)
No	107 (76)
My institution has appropriate regulatory policies in place	
None	16 (11)
Disagree strongly	8 (6)
Disagree somewhat	6 (4)
Unsure	19 (14)
Agree somewhat	48 (34)
Agree strongly	43 (31)
My institution has appropriate ethics guidance in place	
None	11 (8)
Disagree strongly	3 (2)
Disagree somewhat	5 (4)
Unsure	12 (9)
Agree somewhat	44 (31)
Agree strongly	65(46)

HBM, human biological material; DTA, Data Transfer Agreement; MTA, Material Transfer Agreement

Most respondents (64%; 89/140) indicated that they lacked experience in reviewing data-intense protocols that involve data sharing, as up to 50% of all protocols that they reviewed did not relate to data at all, whilst only 14% of respondents (19/140) indicated that more than half of their reviewed protocols related purely to large data sets or big data.

Support systems for REC members

Respondents were asked to indicate the ease of accessing their country's data regulatory body for consultation. Over a third (38%) of respondents (53/140) indicated that they could easily do so, whereas 25% (35/140) disagreed. A portion of respondents (12%; 17/140) indicated that no data regulatory body existed within their country.

A minority of respondents 14% (20/140) indicated that they had received no training on how to review protocols involving data sharing. A fifth (21%) of respondents (30/140) indicated that their institution did not have appropriate regulatory policies on the protection of research data and/or HBMs. Likewise, 14% of respondents (19/140) indicated that their institution did not have appropriate ethics guidance on the protection of research data and/or HBMs (Table 4).

Challenges with data governance

Just over a third (36%) of respondents (51/140) indicated that they faced challenges in their countries regarding the development of

legal frameworks or guidance for research data protection. Only 59% of respondents (82/140) reported having current national laws on data protection. The reasons provided were based on poor resources available within these countries, coupled with a lack of capacity to focus on the development of legislation:

The lack of law is the main challenge to be recorded in SSA. [Country 1]

Specific guidance/law for research data protection is not developed at country level. Laws and [the] Constitution address issues related to data protection in fragmented ways. [Country 2]

Respondents raised a lack of adequately trained legal and ethical experts as another challenge:

The legal experts who develop legal frameworks or guidance for research data protection have not been trained in research ethics. As such, the current legal frameworks for research data protection lack ethical input. Secondly, the current legal frameworks are very restrictive because the regulators are rigid and do not want to move with the signs of the times. [Country 3]

Lack of legal and ethics experts to develop the frameworks...Lack of trained personnel in this field.... [Country 4]

The lack of awareness regarding research ethics and related issues was raised as an issue:

There is a shortage of knowledge amongst clinician practitioners involved in research requiring the implications of the Protection of Personal Information Act. [Country 5]

Respondents also identified the lack of clear DTAs for many countries in SSA as a hindrance to good data governance:

We need to come up with a clear DTA. [Country 6]

Addressing issues related to data in collaborative research. Issues of consent for secondary use of data – use of data for other research not included in the original protocol for which informed consent was provided. [Country 7]

The majority of respondents (66%; 93/140) revealed that they experience some level of difficulty in reviewing data sharing related protocols (Figure 2).

Suggested improvements

Most respondents (71%; 99/140) expressed the view that data sharing for research could be better regulated at their institution. Respondents emphasised a need for the development of institutional policies with clear guidelines for implementation and adequate processes for the follow-up of research protocols. Suggestions around the potential development of DACs within institutions emerged as an idea for the better regulation of data sharing within research.

More than half the respondents (64%; 89/140) indicated that their institutions did not have DACs to handle data-related issues in research. These findings further highlight the need for a DAC as it relates to institutional regulation.

This should start from drafting laws and policies that specifically govern/regulate specimen and data sharing. Research institutions can then draw from these to develop their standard operating procedures or guidelines. External research partners can develop capacity in this area through funding [the] training of IRB members involved in the review of protocols that involve samples and data sharing. [Country 8]

By establishing Data Access Ethics Subcommittees to function under RECs, or better still, provision of training to RECs so that they can play the regulatory role. [Country 7]

Many respondents suggested the development of comprehensive DTAs to improve regulation at a national level. Qualitative responses highlighted the importance of local and international collaboration and the increased need for support to researchers.

The need to raise awareness through education among research stakeholders, including IRB members, researchers, communities, as well as respondents about the benefits and risks of data sharing. This empowerment will encourage research stakeholders to appreciate the need for [the] regulation of samples and data sharing to avoid unethical practices in sample and data sharing like exploitation and harm to individual respondents and communities where the research is conducted. [Country 8]

We need to support researchers to understand the bigger value of data and appreciate [the] value of engaging in data agreements with collaborating institution, which business they have been leaving to the regulator. [Country 9]

Discussion

Historically, RECs have been tasked with reviewing classic clinical trials and other research protocols with limited data sets.²⁴ Robust governance frameworks exist globally and in SSA to guide this type of research review.²⁵ Likewise, a reasonable amount of capacity development has occurred in research ethics review in SSA.²⁵ Big data have raised new ethics and legal challenges²⁶, and our results provide a broad overview of these challenges in SSA. To our knowledge, this is the first empirical survey in SSA in which awareness and perspectives of REC members have been explored specifically as they relate to the review of data-intensive research.

There are governance challenges relating to data protection in research as not all countries in SSA have a legal framework to regulate the use of big data in research. Instead, there is a spectrum of legal regulation, ranging from the strict, comprehensive protection of data to no legal frameworks at all.²⁷⁻²⁹ Likewise, research ethics policies and guidelines suffer the same level of variability across the subcontinent where big data are concerned.²⁵

Our study confirms this variability as knowledge and awareness of legislative frameworks and ethics guidance in SSA vary considerably. Only 58% of the REC members surveyed indicated that laws existed at a national level, with the remainder indicating no knowledge or uncertainty about the existence of such laws. More specifically, a quarter (24%) of REC members were uncertain about whether such frameworks existed within their respective countries or institutions.

Most concerning is the apparent lack of legislative frameworks for the cross-border transfer of big data on the subcontinent and out of Africa to other parts of the world. This is important because of the historical concern with data and samples leaving SSA in an unregulated manner, which raises concerns about exploitative research practices.³⁰⁻³² Although just under two-thirds of respondents were unaware of laws relating to data-intensive research, only half were aware of laws relating to the cross-border transfer of data. This suggests that research data may be crossing borders without agreements or export permits in place. This is supported by Labuschaigne et al.³³ who reported that HBMs may be leaving South Africa without export permits or MTAs during collaborative research. Mwaka and Munabi³⁴, who undertook a similar study on perceptions and experiences on the transfer of HBMs in international collaborative research in Uganda, reported that the development of an MTA and its implementation lacked transparency.

This concern is reflected at a more granular level as knowledge or awareness of DTAs and DACs demonstrate. Our findings reflect this, as 13% of respondents indicated that some countries and/or institutions do not have DTAs or MTAs in place to regulate the national or trans-border sharing of data. While MTAs were more common than DTAs, a fifth of the respondents were not even certain whether such transfer agreements existed within their affiliated institutions. Notably, although our findings indicate the absence of DTAs or MTAs at some institutions within SSA, most respondents (74%) indicated that their RECs were still responsible for reviewing these legal documents together with data sharing-related research protocols when required. This raises concern about the quality of review being conducted on the DTAs and MTAs submitted to RECs. Respondents perceived the development of comprehensive DTAs focused on safeguarding the privacy, anonymity and confidentiality of research participants as an effective resolution. Respondents emphasised that these DTAs should be stringent, with importance placed on institutions instigating mechanisms to improve regulatory compliance. Suggestions included consultation with legal experts in the development of new DTAs, or improvements to current DTAs to ensure that they are aligned to existing laws or regulations. The implementation of access control systems that concentrate on standard criteria for data use and propositions may reduce the likelihood of data misuse, and may legally complement data transfer across borders.

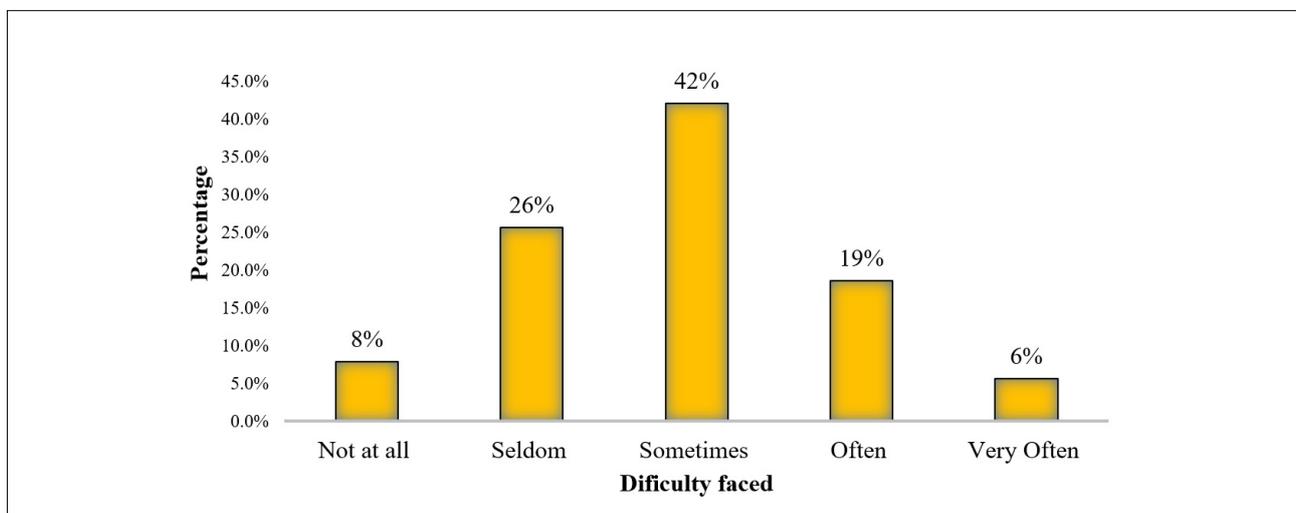


Figure 2: Difficulty in reviewing data-sharing protocols.

Some respondents were of the view that their country's laws were fragmented and consequently exacerbated ethical challenges, thus needing to be harmonised. This was echoed in the responses indicating that data sharing for research could be better regulated both within their institutions (70%) and nationally (71%). Suggestions to develop policies with clear frameworks or stringent standard operating procedures on data sharing emerged, along with improving awareness and access to adequate training on protocol review, data sharing, processing and protection. Likewise, over a third of respondents were not aware of the restrictions placed on the trans-border flow of research data at their institutions.

Many challenges exist in data governance in SSA. The lack of legal and ethics expertise within RECs was recognised as a challenge in adequately reviewing research protocols that related to big data, research transfer agreements and in developing frameworks and policies. Some respondents reported that their institutions do not have ethics (11%) and regulatory (8%) guidance in place for the protection of research data or HBMs, whilst others reported being unsure about whether such ethics (14%) and regulatory (9%) guidance were utilised within their institutions. These findings are comparable with the systematic review conducted by Barchi and Little²⁸, who found that 29 of the 49 SSA countries (59%) had some form of national ethics guidance. Barchi and Little concluded that SSA countries that still lacked regulatory guidance on research data or HBMs would require extensive health-system strengthening in ethics governance before they could be fully engaged in the modern research enterprise.²⁸

Respondents reported the development of adequate legal frameworks or ethics guidance and policies for research data protection within their respective countries as a pressing challenge. A lack of resources was identified as a common reason for this as respondents expressed an increased need for resources, such as training, to efficiently develop and maintain legislative frameworks for data protection in SSA.

Although some of the epistemic gaps presented with RECs could be addressed, some of the committees' responsibilities may be seen as falling outside their mandate and scope of function. This drew attention to the question of who should review such documents when an epistemological challenge exists amongst RECs. Some authors have argued that such responsibility is incompatible with RECs' legislative oversight role and that a legal body is better suited to review such legal documents.¹¹

The current lack of training available in the field of data science for REC members to better handle the ethical, legal and social implications of big data-related research highlights the need to proactively educate and train²⁶ SSA research-based institutions to foster and empower the formation of DACs^{13,35}. While most respondents confirmed that their institutions lacked DACs to handle data-related issues in research, such committees could play a significant role in the data governance ecosystem.^{13,35} The suggestion to form institutional DACs emerged from our study results; however, respondents also indicated that difficulty may be encountered in establishing these committees with members of sufficient and diverse knowledge, skills and experience.

Training needs were evident across the subcontinent. REC members recognised a deficit in their experience and expertise pertaining to the review of research protocols involving big data and related research transfer agreements. This is evident in the large cohort of respondents (64%) that were not often exposed to research protocols that related purely to large data sets or big data as they clearly indicated that the bulk of all research protocols reviewed did not relate to data sharing at all. This finding was further strengthened by the third (32%) of respondents in our study who explicitly stated that they had not received any training on reviewing protocols involving data use and data sharing. Interestingly, 23% of respondents expressed uncertainty on whether they engage with data sharing related research protocols as a result of not entirely understanding what data sharing and big data essentially encompass. This training deficit is not unique to SSA. Ferretti et al.² found that REC members in Switzerland faced similar challenges in adequately reviewing protocols involving big data research due to an existing lack of expertise and experience in the field.^{2,36} In Australia, Pysar et al.¹⁵ revealed that genomic confidence scores in reviewing related research protocols were low amongst REC members

that were less experienced, and had less exposure and training in the field. Hence, most participants (76%) in this study indicated that non-genetics experts that serve on RECs require additional training and/or resources on big data research. Equipping RECs with basic epistemological advantages, in the form of skills and knowledge in big data, would allow them to better fulfil their roles in effectively reviewing data-sharing protocols.

Pisa et al.³⁷ proposed addressing funding issues, strengthening data management systems, providing training and conducting workshops to strengthen regulatory capacity. This will reduce and mitigate instances of data exploitation or harm encountered by research participants and data subjects.

Study limitations

A notable limitation to be acknowledged when interpreting the results of this study is the predominance of responses from some SSA countries compared to other countries (indicated in Figure 1). This may be due to a higher number of RECs in these countries, more active research sites and the fact that it was easier to locate active email contacts from representatives of these SSA countries. These findings were also from a relatively small survey. Potential participants without reliable internet access may have been unintentionally excluded from participation given the internet-based nature of the survey. Because these results were confined to the SSA context, and 15 of the SSA countries did not participate in our survey, we may not have been able to represent the entire continuum of variability present within the SSA region. However, given the absence of empirical studies on the awareness and perspectives of REC members in SSA, these limitations do not pose a major threat to our survey's exploratory aim. Our qualitative research may address some of these limitations and will be published separately.

Overall, our highest number of survey responses was obtained from the Democratic Republic of the Congo, Kenya, Mozambique, Nigeria, South Africa and Uganda. This may be because most of these countries (South Africa, Nigeria, Kenya and Uganda)³⁸ are ranked as the most research-intense countries in SSA by research output in the fields of public health, and environmental and occupational health³⁹⁻⁴¹. The increased research activities in these SSA countries may be associated with increased cross-border data transfer.

South Africa and Kenya are the most stringent in their data export protection. For data to be transferred out of these countries, the data transfer must be purposeful, consent must be obtained from data subjects, and the data processor must verify to the data commissioner that the third-party recipient's jurisdiction is bound by appropriate safeguards for the security and protection of the data.⁴² Yet, our results did not entirely reflect this, as not all responses from Kenya appeared to be in agreement, indicating a divide. Likewise, a divide was observed in the aggregated results from Nigeria, although the country is very research active. This may be because the country's moderately rigid data export protection does not require third-party recipients of data to be bound by adequate data protection laws or agreements in cases where consent is acquired, or where the transfer meets an exception.^{29,38} For South Africa, the highest-ranked SSA country by research output in public health, and environmental and occupational health³⁸, our results reveal consensus amongst respondents regarding cross-border data transfers, which may be due to awareness of POPIA^{29,43}.

Conclusion

In this study, we intended to provide a broad overview of REC members' awareness and perceptions on data governance in SSA and related legal and ethical challenges. Our results uncovered valuable insights and offer a novel contribution to the empirical literature in SSA on big data. Our findings indicate variability in data governance and regulation in SSA, as well as variability in REC members' perceptions of the adequacy of their national laws and institutional policies. Suboptimal awareness of the existence of data protection laws or the lack thereof amongst REC members in the sample was concerning. This will impact negatively on how data-intense protocols are reviewed. There is a unanimous expressed need for the training of REC members on the African continent. Established RECs across SSA would benefit from the reformation of practices and oversight mechanisms, expertise and regulations to better cater for the

big data research context. Transparent, robust and standardised data governance may promote shared ethical values to conduct research with big data on the subcontinent. Data governance within SSA continues to be inadequately supported by legislative and enforcement frameworks.

Acknowledgements

Research reported in this publication was supported by the US National Institute of Mental Health of the US National Institutes of Health under award number U01MH127704. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We thank all respondents for their invaluable time in participating in the study. We also thank Ms Qunita Brown (MA) for her assistance in analysing qualitative responses from the questionnaire. We furthermore acknowledge the following individuals for their guidance during the early stages of the project: Aneeka Domingo, Theresa Burgess, Shenika Singh, Meagan Jacobs-Alfred, Sharon Kling, Gonasagrie Nair and Emmanuel Obasa.

Competing interests

We have no competing interests to declare.

Authors' contributions

K.M. conceptualised the study, reviewed and edited the survey instrument and protocol, and reviewed and edited the manuscript. N.C. developed and submitted the protocol for ethics committee review, designed the online survey instrument, was involved in participant recruitment, led the development of the manuscript and was responsible for data capturing. S.M.K. undertook the data cleaning and analysis, and contributed to the discussion of the results. T.M.E. provided expert advice on statistical analysis. All authors read and approved the final manuscript.

References

- Borgman CL. The conundrum of sharing research data. *J Am Soc Inf Sci Tec*. 2012;63(6):1059–1078. <https://doi.org/10.1002/asi.22634>
- 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 Ethic*. 2022;17(1–2):129–143. <https://doi.org/10.1177/15562646211053538>
- Walport M, Brest P. Sharing research data to improve public health. *Lancet*. 2011;377(9765):537–539. [https://doi.org/10.1016/S0140-6736\(10\)62234-9](https://doi.org/10.1016/S0140-6736(10)62234-9)
- Kaewkungwal J, Adams P, Sattabongkot J, Lie RK, Wendler D. Issues and challenges associated with data-sharing in LMICs: Perspectives of researchers in Thailand. *Am J Trop Med Hyg*. 2020;103(1):528–536. <https://doi.org/10.4269/ajtmh.19-0651>
- Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine, C, James, A, et al. Data sharing statements for clinical trials: A requirement of the International Committee of Medical Journal Editors. *Ann Intern Med*. 2017;167(1):63–65. <https://doi.org/10.7326/M17-1028>
- Can Panhuis WG, Paul P, Emerson C, Grefenstette, J, Wilder, R, Herbst, AJ, et al. A systematic review of barriers to data sharing in public health. *BMC Public Health*. 2014;14, Art. #1144. <https://doi.org/10.1186/1471-2458-14-1144>
- Alter GC, Vardigan M. Addressing global data sharing challenges. *J Empir Res Hum Res Ethics*. 2015;10(3):317–323. <https://doi.org/10.1177/1556264615591561>
- Staunton C, Tschigg K, Sherman G. Data protection, data management, and data sharing: Stakeholder perspectives on the protection of personal health information in South Africa. *PLoS ONE*. 2021;16(12), e0260341. <https://doi.org/10.1371/journal.pone.0260341>
- Knoppers BM, Harris JR, Budin-Ljøsne I, Dove ES. A human rights approach to an international code of conduct for genomic and clinical data sharing. *Hum Genet*. 2014;133:895–903. <https://doi.org/10.1007/s00439-014-1432-6>
- Chalmers D, Nicol D, Nicolás P, Zeps N. A role for research ethics committees in exchanges of human biospecimens through material transfer agreements. *J Bioeth Inq*. 2014;11:301–306. <https://doi.org/10.1007/s11673-014-9552-1>
- Thaldar DW, Botes M, Nienaber A. South Africa's new standard material transfer agreement: Proposals for improvement and pointers for implementation. *BMC Med Ethics*. 2020;21, Art. #85. <https://doi.org/10.1186/s12910-020-00526-x>
- Polanin JR, Terzian M. A data-sharing agreement helps to increase researchers' willingness to share primary data: Results from a randomized controlled trial. *J Clin Epidemiol*. 2019;106:60–69. <https://doi.org/10.1016/j.jclinepi.2018.10.006>
- Cheah PY, Piasecki J. Data access committees. *BMC Med Ethics*. 2020;21, Art. #12. <https://doi.org/10.1186/s12910-020-0453-z>
- 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*. 2020;15():28–39. <https://doi.org/10.1177/1556264619837088>
- Pysar R, Wallingford CK, Boyle J, Campbell SB, Eckstein L, McWhirter R, et al. Australian human research ethics committee members' confidence in reviewing genomic research applications. *Eur J Hum Genet*. 2021;29:1811–1818. <https://doi.org/10.1038/s41431-021-00951-5>
- Ballantyne A. Adjusting the focus: A public health ethics approach to data research. *Bioethics*. 2019;33(3):357–366. <https://doi.org/10.1111/bioe.12551>
- Odusote A. Data misuse, data theft and data protection in Nigeria: A call for a more robust and more effective legislation. *Beijing Law Rev*. 2021;12(4):1284–1298. <https://doi.org/10.4236/blr.2021.124066>
- Jang-Jaccard J, Nepal S. A survey of emerging threats in cybersecurity. *J Comput Syst Sci*. 2014;80(5):973–993. <https://doi.org/10.1016/j.jcss.2014.02.005>
- Ducato R. Data protection, scientific research, and the role of information. *Comput Law Secur Rev*. 2020;37, Art. #105412. <https://doi.org/10.1016/j.clsr.2020.105412>
- Bubela T, Guebert J, Mishra A. Use and misuse of material transfer agreements: Lessons in proportionality from research, repositories, and litigation. *PLoS Biol*. 2015;13, e1002060. <https://doi.org/10.1371/journal.pbio.1002060>
- Department of Health Republic of South Africa. Ethics in health research principles, processes and structures [document on the Internet]. c2015 [cited 2022 Aug 01]. Available from: [https://www.ru.ac.za/media/rhodesuniversity/content/ethics/documents/nationalguidelines/DOH_\(2015\)_Ethics_in_health_research_Principles_processes_and_structures.pdf](https://www.ru.ac.za/media/rhodesuniversity/content/ethics/documents/nationalguidelines/DOH_(2015)_Ethics_in_health_research_Principles_processes_and_structures.pdf)
- Knight J. The need for improved ethics guidelines in a changing research landscape. *S Afr J Sci*. 2019;115(11/12), Art. #6349. <https://doi.org/10.17159/sajs.2019/6349>
- Naderifar M, Goli H, Ghaljaie F. Snowball sampling: A purposeful method of sampling in qualitative research. *Stride Dev Med Educ*. 2017;14(3), e67670. <https://doi.org/10.5812/sdme.67670>
- Tusino S, Furfaro M. Rethinking the role of research ethics committees in the light of Regulation (EU) No 536/2014 on clinical trials and the COVID-19 pandemic. *Br J Clin Pharmacol*. 2022;88(1):40–46. <https://doi.org/10.1111/bcp.14871>
- Nabyonga-Orem J, Asamani JA, Makanga M. The state of health research governance in Africa: What do we know and how can we improve? *Health Res Policy Syst*. 2021;19, Art. #11. <https://doi.org/10.1186/s12961-020-00676-9>
- 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>
- Townsend B. The lawful sharing of health research data in South Africa and beyond. *Inf Commun Technol Law*. 2022;31(1):17–34. <https://doi.org/10.1080/13600834.2021.1918905>
- Barchi F, Little MT. National ethics guidance in sub-Saharan Africa on the collection and use of human biological specimens: A systematic review. *BMC Med Ethics*. 2016;17, Art. #64. <https://doi.org/10.1186/s12910-016-0146-9>
- Brand D, Singh JA, Nienaber McKay AG, Cengiz N, Moodley K. Data sharing governance in sub-Saharan Africa during public health emergencies: Gaps and guidance. *S Afr J Sci*. 2022;118(11/12), Art. #13892. <https://doi.org/10.17159/sajs.2022/13892>
- Moodley K. Research imperialism resurfaces in South Africa in the midst of the COVID-19 pandemic – this time, via a digital portal. *S Afr Med J*. 2020;110(11):1068–1069. <https://doi.org/10.7196/SAMJ.2020.v110i11.15285>



31. Moodley K, Singh S. "It's all about trust": Reflections of researchers on the complexity and controversy surrounding biobanking in South Africa. *BMC Med Ethics*. 2016;17, Art. #57. <https://doi.org/10.1186/s12910-016-0140-2>
32. Singh S, Moodley K. Stakeholder perspectives on the ethico-legal dimensions of biobanking in South Africa. *BMC Med Ethics*. 2021;22, Art. #84. <https://doi.org/10.1186/s12910-021-00645-z>
33. Labuschaigne M, Dhali A, Mahomed S, Behrens K, Nienaber A, Moodley K, et al. Protecting participants in health research: The South African Material Transfer Agreement. *S Afr Med J*. 2019; 109(5):353–356. <https://doi.org/10.7196/SAMJ.2019.v109i5.13803>
34. Mwaka ES, Munabi IG. Trans-border transfer of human biological materials in collaborative biobanking research: Perceptions and experiences of researchers in Uganda [preprint]. *medRxiv*. 2022. <https://doi.org/10.1101/2022.04.01.22273073>
35. Kaye J, Hawkins N. Data sharing policy design for consortia: Challenges for sustainability. *Genome Med*. 2014;6, Art. #4. <https://doi.org/10.1186/gm523>
36. Ienca M, Ferretti A, Hurst S, Puhon M, Lovis C, Vayena E. Considerations for ethics review of big data health research: A scoping review. *PLoS ONE*. 2018;13(10), e0204937. <https://doi.org/10.1371/journal.pone.0204937>
37. Pisa M, Dixon P, Nwankwo U. Why data protection matters for development: The case for strengthening inclusion and regulatory capacity. Center for Global Development; 2021. Available from: <https://www.cgdev.org/sites/default/files/why-data-protection-matters-for-development.pdf>
38. Scimago Journal and Country Rank. Country rankings: Public health, environment, and occupational health [webpage on the Internet]. No date [cited 2022 Aug 01]. Available from: <https://www.scimagojr.com/countryrank.php?region=Africa&category=2739>
39. Lucas-Dominguez R, Alonso-Arroyo A, Vidal-Infer A, Aleixandre-Benavent R. The sharing of research data facing the COVID-19 pandemic. *Scientometrics*. 2021;126:4975–4990. <https://doi.org/10.1007/s11192-021-03971-6>
40. Moorthy V, Henao Restrepo AM, Preziosi M-P, Swaminathan S. Data sharing for novel coronavirus (COVID-19). *Bull World Health Organ*. 2020;98:150. <https://doi.org/10.2471/BLT.20.251561>
41. Capocasa M, Anagnostou P, Bisol GD. A light in the dark: Open access to medical literature and the COVID-19 pandemic. *Inf Res*. 2022;27(2), Art. #929. <https://doi.org/10.47989/irpaper929>
42. Suominen K, Vambell E. Alliance for e-trade development: Toward an African data transfer regime to enable MSMEs' cross-border ecommerce [document on the Internet]. c2021 [cited 2022 May 06]. Available from: https://www.allianceforetradedevelopment.org/_files/ugd/478c1a_72021e35a826441db0723642a79e65e5.pdf
43. Academy of Science of South Africa (ASSAf). POPIA Code of Conduct for Research [document on the Internet]. c2021 [cited 2022 May 03]. Available from: https://www.assaf.org.za/files/2020/POPIA%20CoC%20Research_Conceptnote_Letter%20003.pdf