PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Diverse experts' perspectives on ethical issues of utilizing machine learning to predict HIV/AIDS risk in Sub-Saharan Africa: A modified Delphi study
AUTHORS	Nichol, Ariadne; Bendavid, Eran; Mutenherwa, Farirai; Patel, Chirag; Cho, Mildred

VERSION 1 – REVIEW

REVIEWER	Nepogodiev, Dmitri University of Birmingham
REVIEW RETURNED	29-Apr-2021

GENERAL COMMENTS	Disclaimer -I do not have experience of using DHS data. This is an interesting article but I found its structure difficult to follow. The authors have presented their three stage process in detail and whilst this important for explaining the methodology, it means the results are somewhat jumbled and difficult to navigate. I suggest either 1. Move all methods to the methods section, so the results can flow more seamlessly, or 2. Simplify the methods to present the initial scenario and briefly describe how this evolved (reporting this in the methods section), but move all the other methods information to a supplement For results, rather than describing in details how the three stages proceeded (again detail can be in supplement) it would be more useful to present an overall synthesis of the findings from the study. It may be useful to break this down in to themes specific to DHS/ "big data" and the more generic research ethics themes that are not specific to DHS. There also seems to be two separate strands, issues relating to biobank/DHS use and specific issues relating to predictive tools. It would be useful to separate them out so each can be considered in turn as they seem to me to be different issues of how raw research data is used by researchers, and how research findings are used by public/ policy makers. A lot of information is presented but much of it is cursory making it difficult to understand what it really means. Page 10, Line 161 where the ethical issues are listed, a note should be added that the full item descriptions for these are in Table 1. However, even in Table 1 some of the items are not entirely clear.

	>Potential to stigmatize identifiable groups or populations >> Who is the person potentially stigmatising populations - the researcher or someone else? >Validity of big data analytic tools >> This sounds very broad, validity to what? What is the specific ethical issue? >Potential bias introduced by big data analytic tools >> again very broad, what sort of biases do you have in mind? Another example: "history of human rights abuses, lack of trust in government, misuse of research findings" - it would be helpful to elaborate on this to explain why these issues are important. Why does lack of trust in government matter - is it because they can access the raw data or the production tool? What are the specific concerns about this? How can research findings be misused? Another example: "There was broad agreement on the need for community representatives to have input on how risk factors are described in publications, but less so for input by public health officials" - I'm not sure what you mean by "how risk factors are described"? Do you mean what variables are selected? Further details is needed. Why is PH official input not needed here? Another example: "risks of identification and stigma, there was support for some limitations on reporting (i.e. reporting overall performance of predictive models rather than individual risk factors)" - again not sure what you mean here. What do you mean by identification - do you mean identifying who subject X is in the dataset, or is this about applying the research findings outside of the study? Why would reporting univariable results for association of X and Y lead to identification? What about the ethics of reporting methodology/results... is it appropriate to just present a random model and AUC with no idea where this came from? Some brief summary of what data DHS does/ does not include would be useful for those not familiar. If it is publicly accessible how does ethical review fit in (surely it is redundant once data is public?). Perhaps the data should not be public and there should be a review process to release it? Other points 1. How were the survey invitees originally identified/selected? 2. How was the original scenario designed and who by? 3. How was the qualitative analysis actually done? How were responses coded and themes selected? Who did this? Did you follow reporting guidelines? 4. Would this benefit from more SSA researcher input in the writing group?

REVIEWER	Mwale, M University of Malawi College of Medicine, Public Health
REVIEW RETURNED	09-May-2021

GENERAL COMMENTS	The paper discusses a critical issue concerning public health research specifically SSA HIV and AIDS DHS analytics related ethical issues. However a few areas need to be revisited or clarified before the paper can be published:

	Introduction . In presenting their background to the study, author/s need to contextualize ethical concerns within the context of other global / international conventions or guidelines for research ethics such as the Helsinki Declaration and Charter of Fundamental Rights of the European Union so as to put the study within the perspective of global protocol. Methods . In their sampling procedures perhaps the author/s could best specify the inclusion and exclusion criteria for experts as this is not clear. Results 1. As has been highlighted for the background section the results section could also begin by highlighting any ethical concerns (if any), gaps and anomalies on he use of big data from DHS perhaps raised elsewhere on a global context apart from SSA notwithstanding best-practices on the same. This balanced perspective could help minimize an ethnocentric appraisal but rather present an objective analysis of ethical concerns that arise without a semblance of prejudice or racial bias. 2. In presenting results for survey i and 2 authors need to specify objectively the potential sources of stigma and discrimination to justify the concern and perhaps as highlighted by the respondents. Discussion .I note the use of some statements without proper citation where such is necessary for example: line 245, 'we demonstrated that issues of data privacy, stigma and discrimination, which are *well documented* concerns (no citation)

REVIEWER	Rennie, Stuart UNC School of Medicine Charlotte Campus
REVIEW RETURNED	20-May-2021

GENERAL COMMENTS	This is a thoughtfully designed and articulately presented study of expert perspectives on big data approaches to predicting HIV/AIDS risk in sub-Saharan Africa. I only have a number of minor comments and requests for clarification. The title may not be sufficiently descriptive, as it makes no reference to machine learning or other big data approaches. This may hinder interested readers coming across the paper in the event it is published. On page 6, line 116, it states that consensus was not a goal of the approach. On the other hand, the abstract states that three rounds of iterative surveys were conducted to identify and resolve areas of disagreement [page 3, lines 59-60]. Can this be reconciled? Just for clarification: the professions of the respondents were not collected (and therefore comparisons between them not possible)

	because of the risk of deductive disclosure? I just ask because anonymous surveys sometimes ask for professions, and being able to make comparisons might have added some interesting insights. The scenarios and follow up questions are excellent, but the reader might want to know a bit more about how they came to be developed. Was there influence from other surveys? What was the process? The text states (page 15, line 259) there was a lack of consensus on centralized versus local review, but unlike the disagreements on data access in the same paragraph, there is no sense given of what the disagreements consisted in. Could something brief be added? Page 15, line 261: it says that there was a lack of consensus on whether research on publicly available or de-identified data was considered exempt from ethics review. But what question is this a response to? Earlier in Box 3, it just says that the IRB considered the US researchers exempt. Is the lack of consensus about whether publicly available or de-identified data SHOULD be exempt from ethics review? Page 18, lines 301-303: "This is perhaps in part ..." This might be better in the discussion section than the results section. This may not be relevant, but I was sometimes confused by the use of 'predictive analytics to predict risk of HIV/AIDS' [page 24, line 417] and the use of predictive analytics to predict HIV status [page 6, line 98]. Are these the same thing, and used interchangeably? Because on one reading, predict risk of HIV/AIDS could refer to whether an individual will acquire HIV/AIDS, while predict HIV status could refer to whether an individual has HIV or not. The rationale for Table 2 is unclear to me: is it meant to represent the ethical considerations/concerns unique to this context (as indicated at page 3, line 55)? Is this exhaustive of what could be considered unique from the dataset?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Dmitri Nepogodiev, University of Birmingham

We appreciate the reviewer's suggestion. The sequential ordering of the three surveys is important to present and structure the results around because the first survey was designed based on the literature, while the second and third iterations were based on the responses to each preceding survey. The importance of the sequential order can be demonstrated in how the respondents gave much more detailed answers about their opinions relating to ethics review after we gave them more information about DHS data collection and ethics review, which we had not done in the first Survey. We added language to emphasize the importance of the sequential order within the Survey 1 results, (page 12, lines 304-310). This is also the reason why we do not wish to split out DHS and more “generic” ethics themes because of the original lack of understanding of DHS data by some of the respondents.

A lot of information is presented but much of it is cursory making it difficult to understand what it really means. Page 10, Line 161 where the ethical issues are listed, a note should be added that the full item descriptions for these are in Table 1. However, even in Table 1 some of the items are not entirely clear.
>Potential to stigmatize identifiable groups or populations >> Who is the person potentially stigmatising populations - the researcher or someone else?
>Validity of big data analytic tools >> This sounds very broad, validity to what? What is the specific ethical issue?
>Potential bias introduced by big data analytic tools >> again very broad, what sort of biases do you have in mind?

We thank the reviewer for the suggestion to add a note that the full item descriptions for the ethical issues are in Table 1. We have now added additional language and the sentence now reads: “Ethical issues included privacy, validity, power disparities, alignment and conflicts of interests, benefit-sharing, stigma, and bias (full item descriptions of the ethical issues can be found in Table 1).” (page 9, lines 208-209) The language in Table 1 cannot change since this was the direct language used in Survey 1 that the respondents filled out.

Another example: "history of human rights abuses, lack of trust in government, misuse of research findings" - it would be helpful to elaborate on this to explain why these issues are important. Why does lack of trust in government matter - is it because they can access the raw data or the production tool? What are the specific concerns about this? How can research findings be misused?

We agree with the reviewer that more elucidation of the issues was needed. We have now added in exemplar quotes to Table 2, highlighting in the respondents’ own words why these issues are important and their more specific concerns. This is the added table text (in the paper it is formatted correctly within Table 2):
History of human rights abuses
“How the researchers protect the privacy of these individuals would be critical considering the gross human rights abuses and poor legal frameworks in certain jurisdictions across Africa.”
Lack of trust in government and potential for misuse of research findings
“The most important ethical consideration would be to ensure that the privacy of the individuals in the dataset is not compromised, and government officials have no way of tracing back individuals in the dataset up to the household level.”
“Trust - Entrusting Ministries/governments could misuse the information - how can this be safeguarded. Information and political use - interventions may be denied where political support is low in some regions. Development of tools which could be abused by authorities or for political reasons.”
HIV-associated characteristics (e.g., homosexuality) that are crimes in some African countries
“Since HIV infection is associated with homosexual behavior which is criminal in many SSA countries, individuals identified in the study may also be in legal jeopardy.”
“How will these researchers ensure that their results will be used for good and not for harmful or discriminatory purposes, especially considering that e.g. same-gender sexual relationships are illegal in many African countries, and that people who engage in them are actively persecuted in many?”
Lack of expertise in big data analysis
“Knowledge and understanding of what is big data - for ministries and for the populations.”
Lack of agency of African researchers and ethicists

“There is lack of expertise in ethics review and monitoring research involving big data.”
“The Americans (and their funders) should be in Africa, training Africans in big data methods and tools.”
Compliance with or lack of country-specific laws and policies
“Consider laws in each region/country as these may differ significantly, or simply not exist in a functional format. Important to understand what local laws are available and what is constitutionally acceptable.”
“…Information may have been deposited on an open source without permission, or in violation of the in-country laws.”

Another example: "There was broad agreement on the need for community representatives to have input on how risk factors are described in publications, but less so for input by public health officials" - I'm not sure what you mean by "how risk factors are described"? Do you mean what variables are selected? Further details is needed. Why is PH official input not needed here?

We thank the reviewer for their comment. When we write “less so for input by public health officials”, we mean that there was less of a consensus among respondents that it is necessary to have public health officials input on how risk factors are described in potential resulting publications. We have now added additional language to clarify: “There was broad agreement on the need for community representatives to have input on how risk factors are described in publications (e.g., if local geographic regions were to be mentioned in publications, community representatives would know whether this could lead to stigmatization against those relevant sub-populations), but there was less consensus as to whether it was necessary to obtain input from public health officials." (page 17, lines 552-557)

Another example: "risks of identification and stigma, there was support for some limitations on reporting (i.e. reporting overall performance of predictive models rather than individual risk factors)" - again not sure what you mean here. What do you mean by identification - do you mean identifying who subject X is in the dataset, or is this about applying the research findings outside of the study? Why would reporting univariable results for association of X and Y lead to identification? What about the ethics of reporting methodology/results... is it appropriate to just present a random model and AUC with no idea where this came from?

This is an important point. Identification of individuals or small groups could happen in a few scenarios. For example, if a very specific association is reported – say risk among widows 15–19-year-olds in a particular district – this could make it possible to identify or come close to identifying individuals. Related, high risk among a small group may increase likelihood of stigmatization in the sense of assumption of risk among individuals as a spillover from group-level association. Reporting of model performance may provide a sense of the power of the method without delving into individual predictors. There is a tradeoff in avoiding reporting of individual risk factors, naturally, but this consideration is nevertheless noted. In the response to the reviewer’s comment, we have now changed the language to read: "risks of identification and stigma, there was support for some limitations on reporting to protect the identity of individuals or small groups (i.e. reporting overall performance of predictive models rather than individual risk factors)". (page 17, line 549)

Some brief summary of what data DHS does/ does not include would be useful for those not familiar. If it is publicly accessible how does ethical review fit in (surely it is redundant once data is public?). Perhaps the data should not be public and there should be a review process to release it?

A fuller description is in the scenario in Fig. 1 of Supplemental Information, which we now refer to, and we have also added to Introduction (page 5, lines 106-109): “Namely, these included the particularly sensitive nature of HIV/AIDS, especially in SSA countries, the granularity of the data (including household wealth, educational history, marital status, and the location of households' villages or neighborhoods)"

Other points
1. How were the survey invitees originally identified/selected?

We have now added this information: "Our multi-disciplinary research team, with backgrounds in bioethics, biomedical informatics, and public health in developing countries, identified 35 experts in informatics (n=10), African public health and HIV/AIDS (n=9) and bioethics of Africa-based studies (n=16) that were known to team members to have expertise in the context of public health or HIV/AIDS in Africa, through searches of the biomedical and ethics literature (again, focusing on public health, HIV/AIDS, and the African context), and by snowball sampling." (page 7, 139-145)

2. How was the original scenario designed and who by?

We had previously included some of this information with the following language: "Survey 1 began with a scenario describing an actual research study funded by the National Institute of Allergy and Infectious Diseases at the US National Institutes of Health (Box 1). The study utilizes large, publicly-available survey cohort data that includes detailed health data and HIV status of millions of survey participants throughout the world, socioeconomic data, and Global Positioning System (GPS) coordinates of randomly displaced neighborhoods by up to 5km to protect privacy."

Now, we have now adjusted the language to make this information more clear (page 8, lines 176-180): "Two research team members (MC and EB) developed the scenario for Survey 1 that was based on an actual research study funded by the National Institute of Allergy and Infectious Diseases at the US National Institutes of Health and conducted by some of the team members (Box 1). The scenario briefly describes aspects of the DHS survey datasets that are used but does not explicitly name them."

Also, in the footnotes under "Contributors", we have indicated that EB and MKC contributed to the conception and design of the study and presentation of the scenario.

3. How was the qualitative analysis actually done? How were responses coded and themes selected? Who did this? Did you follow reporting guidelines?

We have provided the requested information (page 11, lines 249-255): "Responses to open-ended questions were analyzed as qualitative data. Statements were initially coded by one of the research team members (MC) to characterize the types of ethical issues or concerns that were raised, such as stigma, data ownership, or the need for stakeholder engagement. These codes were derived directly from the data. We then identified themes representing the most frequently occurring codes where there was lack of consensus or widely divergent views. SRQR reporting guidelines were used.(13)"

4. Would this benefit from more SSA researcher input in the writing group?

We have an experienced SSA researcher, Dr. Farirai Mutenherwa from the University of KwaZulu-Natal, who is a co-author on the paper and was involved in the drafting and finalization of the manuscript and gave valuable input in the writing group.

Reviewer: 2
Dr. M Mwale, University of Malawi College of Medicine, Mzuzu University Faculty of Education

Introduction

In presenting their background to the study, author/s need to contextualize ethical concerns within the context of other global / international conventions or guidelines for research ethics such as the Helsinki Declaration and Charter of Fundamental Rights of the European Union so as to put the study within the perspective of global protocol.

We thank the reviewer for their comment. The first survey of ethical issues came from the literature – including international literature mentioned by the reviewer, such as the Helsinki Declaration. We have now added additional language to make this clear (page 5, lines 111-114): “While many international regulations, guidelines, and conventions already apply to biomedical research (9–12), we sought to understand whether using new types of predictive analytics on sensitive, publicly available data raised additional issues that warranted special attention by researchers.”

Methods

In their sampling procedures perhaps the author/s could best specify the inclusion and exclusion criteria for experts as this is not clear.

We thank the reviewer for the suggestion. We added more detail (page 7) to the sampling criteria as described above, and added that panelists were English-speaking.

Results

1. As has been highlighted for the background section the results section could also begin by highlighting any ethical concerns (if any), gaps and anomalies on he use of big data from DHS perhaps raised elsewhere on a global context apart from SSA notwithstanding best-practices on the same. This balanced perspective could help minimize an ethnocentric appraisal but rather present an objective analysis of ethical concerns that arise without a semblance of prejudice or racial bias.

We agree with the reviewer. In the results section, we added that concerns raised by respondents were not unique to SSA (page 12).

2. In presenting results for survey i and 2 authors need to specify objectively the potential sources of stigma and discrimination to justify the concern and perhaps as highlighted by the respondents.

We fully agree with the reviewer. We have now added more exemplar quotes in Table 2 by our esteemed respondents that highlight the potential sources of stigmatization and discrimination. As a few examples:

“Since HIV infection is associated with homosexual behavior which is criminal in many SSA countries, individuals identified in the study may also be in legal jeopardy.”
“How will these researchers ensure that their results will be used for good and not for harmful or discriminatory purposes, especially considering that e.g. same-gender sexual relationships are illegal in many African countries, and that people who engage in them are actively persecuted in many?”

We have also added in the text on page 12, lines 291-300, another example from a respondent: “Open-ended responses were exceptionally rich, and reflected issues of re-identification, stigma, discrimination against individuals, families, or geographically defined and/or socially defined groups, especially pointing to the possibility of linking to HIV risk. These responses were consistent with the importance accorded these issues in the responses to the closed-ended questions which were asked later in the survey. As an example of stigma and discrimination, one respondent stated: “Perhaps the most concerning is the possibility of developing models that are based on source codes that could potentially stigmatize people, who will be labeled as 'at risk' individuals. Stigma is one of the most harmful conditions in HIV care today, and effective interventions are very hard to develop.”

Discussion
I note the use of some statements without proper citation where such is necessary for example: line 245, 'we demonstrated that issues of data privacy, stigma and discrimination, which are *well documented* concerns (no citation)

We thank the reviewer for catching this point. We have now added the following three citations to the sentence:

Vayena E, Madoff L. Navigating the Ethics of Big Data in Public Health. In: Mastroianni AC, Kahn JP, Kass NE, eds. The Oxford Handbook of Public Health Ethics. Oxford University Press; 2019:353-367. doi:10.1093/oxfordhb/9780190245191.013.31

Enserink M, Chin G. The end of privacy. Science (80-). 2015;347(6221):490-491. doi:10.1126/science.347.6221.490

Beck EJ, Gill W, De Lay PR. Protecting the confidentiality and security of personal health information in low- and middle-income countries in the era of SDGs and Big Data. Glob Health Action. 2016;9(1):32089. doi:10.3402/gha.v9.32089

Reviewer: 3
Dr. Stuart Rennie, UNC School of Medicine Charlotte Campus

We are pleased that the reviewer found the study to be “thoughtfully designed and articulately presented”, and we appreciate their constructive minor edits.

The title may not be sufficiently descriptive, as it makes no reference to machine learning or other big data approaches. This may hinder interested readers coming across the paper in the event it is published.

We agree with the reviewer’s recommendation to reference machine learning or big data in the title. We have now revised the title to be: “Diverse experts’ perspectives on ethical issues of utilizing machine learning to predict HIV/AIDS risk in Sub-Saharan Africa: A modified Delphi study”.

On page 6, line 116, it states that consensus was not a goal of the approach. On the other hand, the abstract states that three rounds of iterative surveys were conducted to identify and resolve areas of disagreement [page 3, lines 59-60]. Can this be reconciled?

We thank the reviewer for catching this point. The reviewer is correct; consensus was not a goal of the approach. We have deleted “resolve” and the language now reads: “… three rounds of iterative surveys to identify areas of disagreement” (page 3, 59-62).

Just for clarification: the professions of the respondents were not collected (and therefore comparisons between them not possible) because of the risk of deductive disclosure? I just ask because anonymous surveys sometimes ask for professions, and being able to make comparisons might have added some interesting insights.

Yes, the reviewer is correct. We wanted to take every effort to preserve the anonymity of respondents, especially given the unique subject expertise involved, so we did not ask for professions of the respondents on the surveys.

The scenarios and follow up questions are excellent, but the reader might want to know a bit more about how they came to be developed. Was there influence from other surveys? What was the process?

Addressed in part above, but also added more about the rationale for survey development (pages 7-8, lines 161-174) “The initial survey was designed to capture a wide range of ethical issues, including those that might not have been already identified in the literature using broad open-ended questions, as well as to assess the perceived importance of previously-raised concerns. Responses were then analyzed to identify areas that were most frequently identified as important but where there was also disagreement about what to do. Subsequent survey questions were developed to identify how experts would prioritize values or make tradeoffs between conflicting values to address ethical issues.”

The text states (page 15, line 259) there was a lack of consensus on centralized versus local review, but unlike the disagreements on data access in the same paragraph, there is no sense given of what the disagreements consisted in. Could something brief be added?

We agree with the reviewer and have now highlighted what the disagreements consisted of: “There was a lack of consensus on the adequacy of centralized versus local ethics review and whether research on publicly-available or de-identified data was considered exempt from ethics review. Some respondents felt the centralized and local ethics review of the DHS surveys presented in the scenario would be adequate and the secondary data analysis of de-identified data would be exempt. However, one respondent articulated a differing view: “Ethics review from the regional and national bodies will be necessary… National ethics committee may be able to instill confidence that there is some oversight. Also any community and national level concerns may then be addressed.” Another respondent disagreed that research on de-identified data should be considered exempt and believed this protocol should “be reviewed (expedited review) by an IRB (ideally based in SSA)”.” (page 15, lines 399-407)

Page 15, line 261: it says that there was a lack of consensus on whether research on publicly available or de-identified data was considered exempt from ethics review. But what question is this a response to? Earlier in Box 3, it just says that the IRB considered the US researchers exempt. Is the lack of consensus about whether publicly available or de-identified data SHOULD be exempt from ethics review?

Yes, the reviewer is correct. There was disagreement expressed within the responses about whether publicly available, de-identified data should be exempt from ethics review. We added in a direct quote to help highlight this point: “Another respondent disagreed that research on de-identified data should be considered exempt and believed this protocol should “be reviewed (expedited review) by an IRB (ideally based in SSA)”.” (page 15, lines 405-407)

Page 18, lines 301-303: "This is perhaps in part ..." This might be better in the discussion section than the results section.

We agree with the reviewer and have now removed the sentence.

This may not be relevant, but I was sometimes confused by the use of 'predictive analytics to predict risk of HIV/AIDS' [page 24, line 417] and the use of predictive analytics to predict HIV status [page 6, line 98]. Are these the same thing, and used interchangeably? Because on one reading, predict risk of HIV/AIDS could refer to whether an individual will acquire HIV/AIDS, while predict HIV status could refer to whether an individual has HIV or not.

We thank the reviewer for bringing the inconsistent language to our attention. We now use “predict risk of HIV/AIDS” consistently throughout the paper.

The rationale for Table 2 is unclear to me: is it meant to represent the ethical considerations/concerns unique to this context (as indicated at page 3, line 55)? Is this exhaustive of what could be considered unique from the dataset?

The Table 2 is meant to provide exemplar quotes of the unique contextual factors that amplify ethical concerns in this context. We now have added in exemplar quotes of all contextual factors listed on page 13, lines 333-339, which were seen in the Survey 1 responses. Please see our response to Reviewer 1 for examples of text we provided in the table.

VERSION 2 – REVIEW

REVIEWER	Mwale, M University of Malawi College of Medicine, Public Health
REVIEW RETURNED	27-Jun-2021

GENERAL COMMENTS	Can be accepted as revised.

REVIEWER	Rennie, Stuart UNC School of Medicine Charlotte Campus
REVIEW RETURNED	12-Jul-2021

GENERAL COMMENTS	I would like to thank the authors for responding adequately to all my minor comments on the previous version of the manuscript. I just have two very minor edits to add: Line 70: there should be a period after ‘surveys’, followed by a new sentence starting with ‘However.’ Common Rule should (I think) be capitalized, and perhaps 'US Common Rule' makes it clearer for an international audience.