Author’s response to reviews

Title: Taking patient involvement seriously: A critical ethical analysis of participatory approaches in data-intensive medical research

Authors:

Katharina Beier (katharina.beier@medizin.uni-goettingen.de)

Mark Schweda (mark.schweda@medizin.uni-goettingen.de)

Silke Schicktanz (silke.schicktanz@medizin.uni-goettingen.de)

Version: 1 Date: 11 Jan 2019

Author’s response to reviews:

Answers to Editor/Reviewers

Editor Comments:

I agree with the reviewers that this article has potential merit as a debate paper. I would like to emphasize some of their points:

1) the "Background" section of a debate paper should contain most of the background information, such that the actual debate and new thoughts can come shortly after. There is still a lot of material beyond the Background section that would rather belong to this section. As pointed out by a reviewer, it would be unusual for most of the original points of a debate to be made in the conclusion.

Answer: We agree with this recommendation and have restructured the sections of the article accordingly. After a short introduction, which outlines the general topic and aim of the article, the background section has been extended to include explanations of relevant terminology and theoretical concepts as a basis for the following discussion.

In addition, we agree with the editor (as well as with reviewer 3 who made a similar point) that it is unfortunate to make most of the original points in the last section. In order to avoid this impression, we have changed the headline of the last section. In fact, the aim of this section is not merely to provide a concluding summary of former arguments but to derive practical and theoretical consequences for the future application of participatory terminology from the analysis. In order to make this clearer, we edited the discussion carefully. Specifically, where possible, we made additions in the discussion part in order to provide a closer link between the genuine analysis of different uses of participatory language in HBDR and the practical and theoretical implications outlined in the last section.
2) As highlighted by a reviewer, we need clarity on the terms. "Big data" has turned into a catch-all in several papers, and sometimes it's peripheral so we can accept it. However, in this debate paper, what is captured by "big data" is important to the debate. I thus support the idea that the background should clarify the terms, and consequently the application context of the debate. As it stands, several sections in the discussion do not clearly show a relevance to big data in particular. Also note that "predictive modelling" and "algorithm-supported clinical decision making" (also commonly called 'clinical decision-support systems') would need to be clarified.

Answer: We agree with the editor’s recommendation that it is useful to define central terms and concepts. Therefore, we added a definition of big data and outlined its specific relevance for the context of medicine (p.3). Given that we do not take up the issue of predictive modelling and clinical decision-support systems, we deleted these terms as we agree with the reviewer that these would require further explanation.

3) Reflecting on this useful debate, the following could help to provide interesting examples and expand on some points:

- While there is sometimes a propensity to use "any data we can" from participants, that is not a universal tendency. There is also a pushback in some areas on whether we should even ask patients, that is, whether knowing individuals’ characteristics or experiences would have added value. This tension is summarized in the section "Modelling human knowledge" in https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-14-130

Answer: We thank the editor for this comment as it allows us to highlight the specificities of the use of big data in medicine. We added a reflection on the argument made by the editor (p. 3).

4) Related to the above, the authors state that "In order to improve the understanding of diseases and to develop future therapies, the availability of large data sets from a broad range of patients and healthy individuals is an important prerequisite". That may depend on the signal we get from the data. If there is a clear pattern and very little heterogeneity across patients, then 'small data' is enough, and perhaps more cost-efficient when collecting additional data comes at a cost. However, when the signal is weak (e.g. low prevalence of rare diseases) or there is a similar heterogeneity (e.g., conditions with complex etiology such as obesity), then more data may be needed. It may thus be useful to nuance when large data sets are actually needed.

Answer: We thank the editor for this comment. In the revised version of our paper we acknowledge the fact that big data may not be desirable for/applicable to all contexts in a similar manner. In order to strengthen our argument of big data playing a crucial role in medicine, we point out two specific contexts where big data is commonly seen as relevant on p.3
There are cases in which a by-product of individuals' activities is to provide data for health studies. They may do it in exchange for a service (c.f., https://link.springer.com/chapter/10.1007/978-3-642-39173-6_23) or may not be fully aware of it. "What" individuals consent to can thus be problematic.

Answer: We understand the editor’s comment as pointing to the problem of potentially insufficient understanding and thus deficient consent of participants in HDBR. We added a reflection on this aspect at three points of the paper: First, we extended the identification of ethical challenges with regard to HRBD at the beginning of the paper and included the problem of informed consent here (p.3). Second, we make a respective remark on this issue at the end of section A at p. 8 and also include the reference mentioned. Finally, we address this point in connection to 23andMe as this is an illustrative example of participants being not necessarily fully aware of potential future usage of their data for research (see p. 9).

Rik Crutzen (Reviewer 1):

1. The manuscript used the term 'big data.' It would be worthwhile to reflect on and/or reconsider this term in the light of the question: to what extent is this only relevant to 'big' data or does this apply to (medical) data in general?

Answer: We agree with the reviewer’s comment and have incorporated this aspect at p.3. (See also our answer to the editor’s second comment above).

2. It is stated that "the collection and analysis or large data sets gives rise to ethical, legal and social concerns." With regard to legal concerns; it is warranted to link the ethical debate to the General Data Protection Regulation (GDPR), which concerns data from all people in Europe (not only citizens, also tourists and refugees), regardless of where they are stored (also when they are stored in the US, for example). Some issues (within GDPR) that should be incorporated in the ethical debate:

   a. Medical data is mostly sensitive data; warranting stricter regulations regarding data processing.

   b. Scientific research has a 'status aparte.'

   c. Article 22 concerns automated individual decision-making, including profiling, which might be relevant in the context of algorithm-supported clinical decision making.

Answer: We thank the reviewer for pointing out these further ethical and legal challenges that arise in the context of big data. As the genuine focus of our debate article is on the normative role and practical usage of participation, we cannot provide an in-depth analysis of each of these
points. However, we mention them, including the new GDPR, in the introduction and clarify the thematic focus of our paper against this broader ethical background.

3. Anonymization concerns rendering of personal data in such a manner that the data subject is not or no longer identifiable. However, anonymization is often incorrectly used in the vernacular when actually pseudonymization is the appropriate term. For example, Sweeney (2000) found that 87% of the population (248 million people at the time) in the United States could be identified based on their 5-digit zip code, gender, and date of birth. So, a dataset containing such variables (in their raw format) cannot be considered anonymized. It is difficult to draw a hard-and-fast line between pseudonymization and anonymization as it depends on the efforts that need to be taken within a certain setting. How is this related to and relevant for the ethical debate?

Answer: We included the challenges related to anonymization and pseudonymization in the newly inserted overview of ethical challenges in the field of big data at p. 3.

4. Related to part A of the discussion, it is recommended to reflect on participating having the right to obtain from the data controller confirmation as to whether or not personal data concerning him or her are being processed, and where that is the case. Participants do not only have rights regarding _access_ to their data, but can also ask for _erasure_ of their data. To what extent are there ethical concerns here; both from the side of the participant as well as the side of the researchers?

Answer: We mention this point in the newly inserted enumeration of ethical challenges in the field of big data at p. 3. However, as our major concern is with the ethical challenges resulting from (partly unjustified) uses of participatory language, a more extensive analysis on this aspect would distract from the original topic.

5. There have been pleas for full disclosure: making data (as well as syntax, output, and other study materials) available in order to maximize scrutiny, foster accurate replication, and facilitate future data syntheses (e.g., meta-analyses) (Crutzen, Peters, & Abraham, 2012; Peters, Abraham, & Crutzen, 2012 - note: I'm amongst those authors). This is part of a more open research culture, as promoted by the guidelines of the Transparency and Openness Promotion (TOP) Committee at the Center for Open Science (Nosek et al., 2015). To what extent does this have ethical implications?

Answer: As the major focus of our paper is on the role of participation, we cannot discuss this aspect in detail here. However, we agree with the reviewer that the tendency to open data and full transparency does have ethical implications, too. We therefore take up this issue in the revised version at p. 12.
When discussing the 23andMe example, it might be worthwhile to reflect on the ethical implications of informed consent procedures. The description of having consent within the GDPR is "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her." This one-sentence definition contains multiple aspects:

- Freely given implies, for example, that there is no imbalance of power between data controller and data subject and that withdrawing consent should be as easy as providing it (e.g., unticking a box).

- Specific implies, for example, that there should be a purpose specification as well as a specification of retention periods. The purpose specification should describe why the data controller needs the data. The GDPR allows a more broadly formulated purpose specification for scientific research. However, this does not mean that "gaining more insight into human behaviour" could be considered as being specific. More details regarding the study goals need to be provided. When there are multiple study goals, data subjects should agree on processing of personal data for each of these goals separately. With regard to retention periods, the GDPR states that these should be "no longer than strictly necessary."

- Informed implies, for example, that it should be in clear and plain language (i.e., level B1 within the Common European Framework of Reference for Languages); not full of legalese.

- Unambiguous implies, for example, that it is a clear affirmative action (e.g., opt-in instead of opt-out).

All of these aspects have ethical implications; please elaborate on this.

Answer: We understand the reviewer’s comments as expressing a general concern that the ethical and legal issues raised by big-data based research are not sufficiently addressed. Again, we agree with the reviewer that informed consent and the related issues he outlines are important from an ethical perspective. However, as our debate does not intend to provide a comprehensive picture of all ethical issues in the field (which would be beyond the scope of a debate article), but rather to focus on the use of participatory language in this field, we decided to present an extended overview of ethical issues, including the problem of informed consent, in the introduction (p.3) and restrict the genuine analysis and discussion of our paper to the issue of participation.

In addition, we make explicit mention of ethical problems related to informed consent in the context of our analysis of 23andMe and included a reference here, too (p.9). Moreover, we take up this issue also at the end of our analysis in section A where we discuss initiatives that equate participation mainly with individual consent (p.8.)
7. The comments above are meant to be constructive. In general, the content of the manuscript is adequate. However, especially since it is a debate paper, it would be nice to be a bit more thought-provoking (not necessarily provocative, of course). For example; what should the reader take home from the manuscript? What questions should s/he ask him/herself after reading the manuscript? It would be worthwhile to make this more explicit. This might even result in commentaries (which helps in bringing this topic to the attention of more readers).

Answer: We like to thanks the reviewer for this encouragement. We have now inserted at the end of the last section a box with take home messages. They serve as thought-provoking, short statements resulting from our article and illustrate the character of a debating article.

8. Typos:
   a. Page 2: methodlogical methodological
   b. Page 3: developmet development

Answer: We have corrected these typos. Moreover, the revised manuscript has been edited by a native speaker.

References


Answer: We are grateful for being pointing to these references. We have included them at the respective parts of our manuscript.

--------------------------

Laura Schmitt Olabisi (Reviewer 2):
I am not in the biomedical field, nor do I work extensively with big data. I believe I am reviewing this manuscript on the basis of my expertise in participatory and engaged research (so that the authors and editors have context for my comments!) This is a very well-written and thoughtful piece. I have some (minor) suggestions below.

In the abstract 'negligence' is misspelled

Answer: We have corrected this typo. Moreover, the revised manuscript has been edited by a native speaker.

In the discussion of co-production of knowledge, the authors state that this represents the most encompassing realization of citizen science. However, in order for citizens to have true co-ownership over the research, they would need equal access to funding mechanisms and knowledge produced (e.g. scientific journals). So co-producing knowledge alone would not meet this criterion unless the power-sharing and knowledge-sharing between citizens and researchers were truly equal. Obviously, this is extremely rare in practice but it is a vision some projects are working towards (the Flint, Michigan Healthy Flint Research Coordinating Center is one example). The authors give an example of the MIDATA study which might (partially) fulfill this vision also.

Answer: We thank the reviewer for her comment and fully agree with her that citizen science does not stop at the point of knowledge co-production. We therefore revised the respective sentence on p.11 and also added an explanation on this issue in the outlook section.

Relatedly, I am curious about an aspect of the MIDATA study which perhaps deserves more discussion in the manuscript—specifically, the interpretation and use of information. One of the barriers to participation is that scientific findings—even if co-produced by citizens—are often reported and interpreted in a way that is difficult or impossible for nonscientists to understand. Moreover, most citizens do not have access to academic journals. Science communication is therefore a vital part of participation which is not touched on here. I am curious how results of research were reported in the MIDATA study and if the participants found these results intelligible and satisfactory.

Answer: We thank the reviewer for this comment as it shows us that the MIDATA initiative deserves further explanation. MIDATA is not a study but a newly introduced IT infrastructure in Switzerland which can be used for research. We clarified this by shortly discussing one example of a current research project and adding a reference to the website of the initiative (p. 11/12).

Page 11, line 26: I think you can say a bit more here. Can you give a specific example of these biases, and what it would look like to remove them?

Answer: We clarified this point by adding an explanation on what kind of biases we have in mind. On the one hand, if participants can enrol themselves to studies this may lead to unbalanced research cohorts that might not be suitable to answer a specific research question
(p.11). On the other hand, biases may exist due to different preconditions that impact people ability to act as co-decision-makers in research (p.12).

Marcello Ienca (Reviewer 3):

This is an informative and well-written review article, which seems suitable for the Debate section of the journal. The article introduces some useful conceptual distinctions and suggests some basic criteria for ethically justified participatory approaches to BDMR.

A few minor remarks:

- The most original part of the paper is relegated to the "conclusion" section. Ideally, the conclusion is the section where the most important discussion points of the article are summarized, not the one where new ideas are presented. While I understand that these conclusions logically follow from the premises delineated in the previous sections, I would recommend re-organizing the paper's structure or, at least, considering to rename the final section.

Answer: We agree with this recommendation and have restructured the sections of the article accordingly. After a short introduction which outlines the general topic and aim of the article, the background section has been extended to include explanations of relevant terminology and theoretical concepts as basis for the following discussion.

In addition, we agree with the reviewer that it is unfortunate to provide the most original points in the last section. In order to avoid this impression, we have changed the headline of the last section. In fact, the aim of this section is not to merely provide a concluding summary of former arguments but to derive practical and theoretical implications for the future application of participatory terminology from the prior analysis as it seems appropriate for a debate article. In order to clarify this, we edited the discussion carefully. Specifically, where possible, we made additions in the discussion part in order to provide a closer link between the genuine analysis of different uses of participatory language in HBDR and the practical and theoretical implications outlined in the last section.

- Authors have emphasized that, in a big data ecosystem, it is hard to distinguish strictly medical from non-medical data, hence have suggested to use the language of "big data health research" or "health-related big data research" to underscore how non-medical data sources can be used for health-related research purposes. It would be advisable that the authors include a mention to this conceptual discussion in the background section of their manuscript.

Answer: We agree with the reviewer on this issue. In fact, big data-based research in medicine does not only include genuine medical data, but also sociodemographic, life style data etc. We
point out to this in the newly inserted definition of big data (p.3). In addition, we have changed our terminology by speaking of health-related big data research throughout the article.

- The article is full of quotations. While I understand this stylistic choice is aimed at minimizing the risk of misrepresenting the cited papers, this unusual -and, from my perspective, excessive- number of direct quotations makes the paper uneasy to read.

Answer: We understand the reviewer’s concern regarding readability. Therefore, we reduced the amount of direct quotations as far as possible. Furthermore, we use single inverted commas to indicate that these are ‘key ore jargon terms’ used in the debate – and not invented by us. However, as the aim of our paper is to provide a critical analysis of current uses of participatory language, there remain few cases, where we regarded it important to cite the original wording as they display self-portrayals of the respective initiatives or their portrayal in the public (e.g. in the media).

- Background/p.2, 145-50: Since the care.data program was abandoned in July 2016 it would be preferable to use the preterite instead of, respectively, the present and present perfect. E.g. "was introduced", "it aimed at" etc.

Answer: We agree with the reviewer’s suggestion and have revised the text accordingly.

- P.3, 151: "Against this background": ambiguous formulation, please rephrase.

Answer: We have deleted this ambiguous formulation.

- P.4, 116-17: Reference needed.

Answer: We added a reference to support this claim.

- P.7, 115: Do you mean "line of reasoning?"

Answer: The reviewer’s suggestion is correct. We have changed this typo accordingly.