Reviewer’s report

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

Version: 0 Date: 31 Oct 2018

Reviewer: Rik Crutzen

Reviewer’s report:

The manuscript introduces basic conceptual distinctions for the understanding of participation and formulates basic criteria for justified appeals to participatory approaches in data-driven medical research. Suggestions for improvement:

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?

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.

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?
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?

5. There have been pleads 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?

6. 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.
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).

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

References


Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes
Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

Not relevant to this manuscript

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal.