Author's response to reviews

Title: Research across the disciplines: Quality criteria for empirical-ethical studies

Authors:

Marcel Mertz (Mertz.Marcel@mh-hannover.de)
Julia Inthorn (Julia.Inthorn@medizin.uni-goettingen.de)
Günter Renz (guenter.renz@ev-akademie-boll.de)
Lilian G. Rothenberger (geza_r@hotmail.com)
Sabine Salloch (sabine.salloch-s52@rub.de)
Jan Schildmann (jan.schildmann@rub.de)
Sabine Wöhlke (sabine.woehlke@medizin.uni-goettingen.de)
Silke Schicktanz (silke.schicktanz@medizin.uni-goettingen.de)

Version: 4
Date: 10 January 2014

Author's response to reviews: see over
Dear editors,
dear reviewers,

we are very thankful that we can resubmit our revised paper to BMC Medical Ethics. According to the points raised by the reviewers, we substantially revised the method section and the tables of our paper.

We like to clarify that the structural headlines we use “background, approach, discussion” are in accordance to the authors guidelines by the Journal BMC Medical Ethics for this type of paper. Hence, we keep them but agree that they might be a bit misleading for some readers. Otherwise it would be an editor’s decision to change the current option of “discussion” into “result” as reviewer #1 suggested.

We have substantially revised and enlarged the methodology section (now: page 8-14) to describe the whole process in detail. We also added a figure (now figure 1) to illustrate this long process and to make it more understandable. The whole section is now quite long and we fear that this may reduce the readability of the paper as a whole. However, we see the point that transparency about the methodology is consistent with what we suggest as quality criteria. Hence, we suggest alternatively for publication in case of acceptance that we keep the figure within the paper and provide the extended 4 page description of the method (starting with the sentence “Our search and analysis strategy included first (i)...) and ending with the sentence: Each subgroup proposed their phrasing of the questions to the whole author group to achieve consensus on the final phrasing.”) as supplemental material. If editors and reviewer, however, approve the long version, this would be also acceptable for us.

In the following, we answer to the points of the reviewers as follows (text that was substantially revised or added to address the critique is marked in yellow):

1. Method section

Reviewer #1: “The new methodology section (pp. 8-9) only provides information about the discussion process of the group of authors but does not disclose (on an argumentative basis) how, i.e. for what reasons the group arrived to each of the results that it presents.“

Reviewer #2: “Who was part of each subgroup? How was consensus achieved and what kind of consensus was it (ie did everyone have to agree, or just a majority. Was it vote based or solely based on discussion)? Can the process be demonstrated more clearly by presenting the reader (using figures/tables/flow charts/diagrams) how many criteria were identified at each stage, which ones were kept and which ones rejected?”

One explanation beforehand: The word ‘consensus’ is used in its original philosophical meaning: There was no voting (or majority decision), but only when all members agreed or no one rejected we talk about consensus. As this is generally acknowledged the original meaning of the word ‘consensus’ (in contrast to majority decision e.g.) we believe there is no need for extra explanation.

The whole new section reads now:
Mapping landmarks of quality and drafting a road map

To survey specific “landmarks” of quality in EE research, and to draft a corresponding “road map”, our procedure consists of the following main steps (figure 1 provides a graphical overview of the search and analysis strategy the working groups used during the project):

(i) to analyse selected empirical quantitative and qualitative studies as well as theoretical ethics studies about living organ donation (“bottom-up-strategy”) regarding their use of empirical data and ethical concepts, and if they reflected upon that relationship;
(ii) to study, present and critically discuss already established quality criteria for each of the following three branches of relevant criteria, viz. a) empirical/social science research, b) philosophical/normative-ethical research, and c) EE research (“top-down-strategy”);
(iii) to consider, present and critically discuss research ethics criteria for each of the three branches, in the light of our experience in EE research and knowledge of the EE debate;
(iv) to develop a consensus among the authors;
(v) to refine the different branches and reduce complexity for publication; and
(vi) to draft a tentative checklist of questions which operationalises criteria pertinent to EE research.

Our search and analysis strategy included first (i) a bottom-up analysis of 10 publications, dealing explicitly and/or implicitly with ethical and empirical issues of living organ donation. This field was used as a focused case study to allow a comparison of quantitative (n=3 [28,29,30]) as well as qualitative (n=3 [31,32,33]) empirical studies and theoretical ethics publications (n=4 [34,35,36,37]).

This bottom-up detailed analysis revealed that firstly, the relevance of empirical data for ethical analysis is often not made explicit, secondly, empirical studies tend to be crypto-normative in their conclusions (“crypto-normative” for us implies that implicit evaluations and ethical conclusions are made, but the evaluative step is not explicated), and thirdly, theoretical studies often refer to empirical data, but rarely critically reflect the empirical methodology, or these often tend to apply empirical data in a positivistic manner.

For step (ii), we applied a more top-down strategy by summarising existing literature in quality criteria in the three relevant fields: empirical research in social sciences, philosophy/normative ethics research, and EE research. Thus, we composed three subgroups in our Empirical Ethics Working Group (comprising about 13 active members at that time; see endnote (i) for further information about the working group). Each group conducted a review of the methodological literature in the relevant area. We also included explicit recommendations for quality research drafted by scholarly societies (e.g. of psychology or sociology).

Literature was searched with a narrative/selective search strategy [see e.g. 38]. This strategy was developed due to difficulties and inappropriate results when trying a systematic literature search by using specific search terms, given the interdisciplinary nature of our topic. As a consequence we decided to broaden our approach by using literature found via PubMed, Philpapers and Google Scholar, via manual search of scientific journals, as well as via expert opinions generated from the members of our working group and their connections to the respective scientific community.
The subgroup of trained philosophers (MM, with non-authors JD and UM; see Acknowledgments) analysing the criteria within philosophy often had to extract informal and implicitly given criteria (apart from criteria directly related to standards of argument, e.g. logic). The principal subcategories of criteria for philosophical, normative-ethical research are the criteria of good argument (as discussed in informal and formal logic [see e.g. 48,49]), the use of specific philosophical methods (theories, approaches) with their respective quality criteria [e.g. 50], and criteria of good ethical judgement and/or decision-making (as discussed in models and methods of decision-making [e.g. 51] (see figure 2).

The subgroup on quality criteria for empirical research in social sciences (LGR, GR) had to differentiate the literature search of journals and monographs into general criteria (such as adequacy of the research process, transparency, good scientific and ethical conduct) and specific criteria for quantitative methods [e.g. 39], qualitative methods [e.g. 40,41,42,43,44] or mixed-methods approaches [e.g. 45,46,47] (see figure 2).

The subgroup working on quality criteria for EE research (JS, SaS) performed a selective literature review in relevant bioethics journals and books focusing on theoretical and methodological contributions to EE research. While a number of conceptual accounts to EE were identified, the issue of quality standards was only rarely addressed [52,53]. However, parts of the conceptual considerations as identified in the literature could be translated into criteria relevant to EE research [54,55]. In addition, the researchers drew from current EE research addressing ‘end of life issues’ which is performed in an interdisciplinary research group of medical ethicists with a disciplinary background in philosophy, sociology and medicine.

Following these procedures, the findings of each subgroup were discussed within the whole working group. Our next step was to systematise the criteria identified by clustering them into main categories according to their field (empirical, philosophical, EE research), as well as into subcategories (e.g. different categories for qualitative and quantitative empirical research, or different categories for criteria related to logic/argumentation theory and philosophical approaches). The clustering was based on the literature search (inductive strategy), as well as on our own theoretical estimation of aspects relevant for assessing the quality of scientific work (deductive strategy). The summarising process was supported by mind mapping software to track modifications, deletions, additions or re-locations of criteria. The main and sub-categories in this mind map were generated either inductively on the basis of the literature or deductively by own reasoning against the backdrop of scientific experience and theoretical knowledge.

Finally, we derived three overarching standards of scientific research, which were subdivided into formal, cognitive and ethical norms (see below Three peaks that dominate the scenery) based especially on the philosophy of science. The actual quality criteria were then seen as specific expressions of these overarching standards.

In step (ii), issues concerning research ethics already found in the reviewed literature were added to the mind map as further criteria. Additional literature was also reviewed [56,57,58,59].

A consensus round was initiated for step (iii) where each criterion was again critically discussed with a view to identify possible redundancies. Consensus was reached with the results of the argumentation in the discussion, which was most often accompanied by a final, explicit request if there were any dissenting votes regarding the result. Active members of the working group who
were not able to participate at the consensus round (about 3 out of 13 members) had the 
opportunity to show assent or dissent on the basis of the sent draft of the mind map; there was no 
crucial dissent that led to a substantial revision of the mind map.

The next step (iv) consisted of focusing on those criteria that were seen as only specific to and 
coherent with EE research, excluding those relating to broadly empirical research in social sciences 
and philosophy. With this aim in mind, the working group agreed to divide specific EE criteria into 
four domains (see also figure 1), effectively reducing the amount of criteria in the mind map of about 
200 (all branches, research ethics included) to about 50. In these four domains, the formal and 
cognitive norms relevant to all kinds of scientific research were specified and adjusted to the 
particular field of EE research (“primary research question, theoretical framework & methods” and 
“relevance”). Furthermore, the specific interdisciplinary nature of EE research was addressed 
(“interdisciplinary research practice”), and issues of research ethics which are pertinent to EE 
research (“research ethics & scientific ethos”) were considered.

As a result, four new subgroups were established that had to summarise, systematise and elucidate 
the according criteria on the basis of the already found literature of the three aforementioned 
subgroups, as well as propose additional criteria if found necessary. The members of these new 
subgroups also became the authors of the paper at hand and were assigned to the four domains as 
following: a) primary research question, theoretical framework & methods (JI, SiS, SW); b) relevance 
(MM); c) interdisciplinary research practice (JS, SaS); and d) research ethics & scientific ethos (LGR, 
GR).

This work also led to the last step (v), the drafting of a refined list which allows the “road map” below 
to be used as a checklist to guide EE research. For clarity, the criteria included in this list are 
presented in tabular form (see tables 1-4). The tables 1-4 contain each criterion, operationalised into 
questions. We decided that it was heuristically more effective to ask questions rather than to 
consider statements, and to conceive these questions as a pragmatic aid to guide scholars and help 
them reflect on their own research. Nevertheless, the questions that operationalise criteria should 
not be understood as simple “yes/no” queries – instead they should function as reflective and critical 
questions designed to assess certain quality-related aspects of EE research. Each subgroup proposed 
their phrasing of the questions to the whole author group to achieve consensus on the final phrasing.

This checklist idea is not new. It is already well established in other research fields, e.g. in medicine 
for guideline recommendations (GRADE [60]; SIGN [61]), quantitative randomised medical trials 
(CONSORT [62]) and observational epidemiological trials (STROBE [63]), where they are used to check 
evidence and/or the quality of (the reporting of) trials. Although normative or especially ethical 
aspects are rarely explicitly mentioned in these checklists, they include implicit normative items such 
as asking for ethical approval, informed consent, funding or possible sources of bias. Critical appraisal 
is more and more coming up on the agenda of evidence based medicine [64; see also 65].

In analogy, we thought it necessary to render explicit ethical questions that are implicit in EE 
research. Looking for the best fitting form of presentation, our working group came to the consensus 
that we would try to adapt the checklist format, as we thought it will be most helpful in order to 
display the suggested criteria in a clear, feasible way.
Reviewer #1: “We do not know which publications have been included in the literature review, how this selection has been made, why certain other texts obviously have been excluded, and what has been taken from these texts that are already existing. […]“

Reviewer #2: “Detail of the search strategies employed - how does the reader know that the literature consulted was reliable, trustworthy or comprehensive.”

The literature is now in detail documented (see also section 2. Tables/Criteria below). For each criterion the used literature is now cited (in the text as well in the tables). We believe this has created the maximum transparency how these criteria were identified. We therefore had to extend the list of references from formerly 54 to now 104. This should convince reviewers and readers, that the list of references considered was comprehensive, trustworthy and reliable. Of course, part of the literature is in German. However, we believe that in an international discipline such as bioethics, this must be acceptable.

2. Tables/Criteria

Reviewer #1: “The criteria that are presented in the 8 tables are not argumentatively justified. They are just falling out of the blue, presented like the minutes of a brainstorming session. […]The group’s own proposals, which are presented in the ‘Discussion’ section […] are not explained with reference to the literature. Therefore it is difficult, or impossible for the reader to see which criteria are novel proposals by these authors and what has already been proposed by others before.”

Reviewer #1: “Some criteria (e.g. compatibility and relevance in table 2) are interesting. But what is meant with “theoretical compatibility”? The brief explanations on p. 13 remain very sketchy. The same for relevance. Other criteria (e.g. “How do socio-theoretical, medico-theoretical, or other theoretical frameworks fit into the ethical or meta-ethical framework?” in table 2) remain difficult to understand. […]”

To address the reviewers’ concerns about the reliability and understandability of the criteria/questions in each table we have substantially revised all tables. Each question was reformulated in a sense it addresses now only one issue and not several ones. For each question/criterion we provide now one clear practical example what this means for an empirical ethical research. For each criterion we also provide now a reference (to be found within the table and within the text, too). If no particular reference is given, we assume that this was our original contribution. (Of course, this does not exclude the fact that someone else might have had the same idea).

We compiled the formerly 8 tables (this was mainly due to format considerations) into 4 tables to make it more consistent with the 4 working groups we mention in the methodology section.
### Table 1: Criteria related to primary research question and selecting a theoretical framework and corresponding methods

**Reflection on the relationship between empirical and ethical–normative/ethical–descriptive research questions (even if there is more an ethical motivation than a clear ethical research question):**

- Can an explicit distinction be made between the empirical and ethical research questions? (e.g. there is distinction between interviewing patients about their wishes and the ethical weight given to patient’s autonomy) [22,24]

- How dependent is the empirical research question on particular ethical background assumptions? (e.g. justification of the selection of target group for a questionnaire: why do we think their opinion is ethically relevant?) [73]

- How is the ethical research question dependent on empirical or socio-theoretical background assumptions? (e.g. ethical considerations of vulnerability of a particular group such as pregnant women: what are the underlying anthropological or psychological considerations? Are there any hidden gender-related stereotypes?) [66]

- What are the explicit and implicit research interests and motivations of the EE researchers? (e.g. is research with dying patients motivated by curiosity or the moral attempt to empower them? Is the researcher motivated to identify possible conflicts of interest or might the research serve mainly to produce more social acceptance of a technology?) [16,69]

- What kind of epistemic research interest motivates the researcher to combine ethical and empirical research? (e.g. explaining whether the aim is the evaluation of established ethical practice, or of measures taken to improve ethical practice; or the aim is ethical theory-building, norm-construction, or legitimization/critique or a particular practice) [17]

**Development/usage of theoretical frameworks:**

- How can a theoretical framework be developed; what are the main limitations of the chosen theoretical framework? (e.g. premises and limitations of a principle of autonomy, when analysing macro social interactions) [22, see contributions in 67]

- Were potential ambiguities of central concepts considered within the theoretical framework? (e.g. how much is the concept of 'identity' differently used in current philosophy and in sociology when one want to analyse empirically the discourse of “identity changes by neuroenhancement and its ethical
• How does the chosen medico-theoretical framework (e.g. concept of disease/health) fit into the ethical-normative framework? (e.g. does a science-positivistic concept of disease fit into a Kantian or hermeneutic approach of ethics?) [70]

• How does the chosen socio-, cultural- or philosophical framework (e.g. concept of personal identity) fit into an ethical normative framework (e.g. approach to a cosmopolitical ethics of justice?) [71]

Usage of empirical and the ethical methods and their relationship to the theoretical frameworks:

• Are the chosen empirical methods compatible with the combined theoretical framework? (e.g. are interviews with doctors as experts compatible with a liberal, autonomy driven approach that claims to empower patients?)

• What is the advantage of the chosen method in comparison to other available methods? (e.g. why and when to choose a deductive approach in applied ethics to assess ethical problems of a new technology and not an inductive, or hermeneutic one?) [66,68]

• Are the chosen methodological approaches appropriate for the envisaged combined research question? (e.g. does the empirical method of interviewing parents generate results relevant for the ethical question of whether parents should be allowed to influence the genetic make-up of their children?)
Table 2: Criteria related to relevance

**Contribution to (scholarly) ethics (epistemic/scientific relevance):**

- Will the study possibly produce knowledge that could not be generated by relying on traditional disciplinary methodologies? (e.g. overcoming too separated empirical research and separate philosophical discussion) [17,22]

- Does the study aim to increase our knowledge, and if so, with regard to what? (e.g. does the study contribute to a balance between theoretically generated norms and empirically found norms? E.g. Does it revise/improve the impact of ethical guidelines?) [56,57,59,77]

- Does the study aim to give input on an ongoing controversy, or does it provide a new perspective on it? (e.g. clarifying if relatives are able to give substitute judgment for incapacitated patients or not, or e.g. if post-trial access should be compelling on the basis of new evidence of consequences when post-trial access is not given etc.) [15, partly 78]

- Does the study aim to offer substantial arguments for or criticism of established ethical positions? (e.g. is a contribution to theory modification or to a refinement of the application of a theory expected? Are descriptive presuppositions of an ethical position, such as anthropological, sociological or psychological assumptions, criticised?) [1,15,66]

- Does the study aim to contribute to the development or refinement of scientific methods, especially methods of EE research, and if so, how? (e.g. pilot testing of a jointly developed research instrument, identifying the need of developing new or refined forms of interactions between researchers) [6,59]

- Does the study aim to offer potential for innovation, and if so, what kind of innovation? (e.g. is it a contribution to theory-building expected? Will the study generate new instruments for ethical decision-making?) [6,79]

- Does the study aim to contribute to another scientific and/or ethical discourse? (e.g. does it contribute to social sciences discourses?) [partly 1]

- Does the study clearly states to whom it is addressed, and who will benefit from its results? (e.g. are the addresses and/or beneficiaries physicians, nurses, social scientists, ethicists or especially empirical ethicists? Are policy-makers or persons in a management position addressed? [59]

**Contribution to ethical practice (societal/practical relevance):**

- Does the study aim to improve ethical decision-making? (e.g. will it produce evidence that was absent,
or will it give guidance regarding specification of established and accepted norms or regarding the interpretation of institutional or legal rules?) [6,16,77, partly 18]

- Does the study aim to raise awareness (among actors, institutions or society) of particular ethical problems? (e.g. does the study identify new ethical problems, or does it highlight specific aspects of a known ethical problem that was not yet addressed sufficiently in practice?) [20,80]

- Does the study aim to lead to a shift in structures and/or decision-making processes (in relevant institutions)? (e.g. establishing new guidelines or building new forms of committees for ethical review) [79, partly 24]

- Does the study aim to establish minimum ethical standards (in relevant institutions or professions)? (e.g. creating new informed consent procedures for specific patient groups?)

- Does the study aim to contribute to or stimulate public debate? (e.g. about physician-assisted suicide, rationing in health care, public health funds etc.)

- Does the study aim to contribute to or stimulate a process of legislation? (e.g. proposing alteration of legal norms)

- Does the study aim to articulate the need for reforms (in a certain institution or system of society)? (e.g. by evaluating current practices.) [6]

- Does the study aim to articulate new ethically pertinent ecological or economic problems? (e.g. costs related to a broad implementation of the use of social robots in elderly care) [15]
<table>
<thead>
<tr>
<th>Table 3: Criteria related to interdisciplinary research practice</th>
</tr>
</thead>
</table>

**Research drafting:**
- What form of interdisciplinary collaboration serves the needs of an EE study? (e.g. how strong and how often should collaborators interact? Is it necessary to have face-to-face-meetings? Who has to be involved in which step of the research? Is there reflection on the potentials and the limitations of the kind of collaboration?) [14]
- How can the participating researchers be adequately selected? (e.g. Which disciplines/methods are actually needed?)
- How can an appropriate task schedule be developed? (e.g. at which point in time is empirical data to be gathered)
- Which agreements must be reached with regard to interdisciplinary communication? (e.g. consideration of terms used, explanation of professional jargon and development of a “common language”) [3]
- How can competencies and responsibilities be reasonably distributed within the research team? (e.g. despite their varying competencies, will all the interdisciplinary researchers remain actively involved in the research process? Who is accountable for what?)
- How can research questions be developed jointly? (e.g. regarding different interests and disciplinary perspectives, or regarding the goal of the study)
- How can the literature search be carried out? (e.g. having to acknowledge empirical-ethical studies from one’s own thematic field as well as to acknowledge both empirical and ethical work from different disciplines in diverse types of publication)

**Data gathering:**
- How is the joint development or modification of a research instrument carried out? (e.g. Is there a process that allows for dissent and argument in developing or modifying a research instrument?)
- Is there normative-ethical reflection on the empirical research process? (e.g. can implicit normativity be revealed that is related to a theoretical background (“social constructivism”)?)
- Is there a mutually critical appraisal by normative and empirical sciences with regard to data gathering? (e.g. what constitutes “good” data for the EE study) [54]

**Data analysis and conclusions:**
• How do normative and empirical aspects interrelate with regard to analysis and deliberation?
  o Is the analysis of the empirical data influenced by normative theories, concepts, or standpoints? (e.g. by a specific account of patient autonomy)
  o Is the normative deliberation influenced by the requirements of the empirical data analysis? (e.g. by standardization of data) [83]
• How do normative and empirical aspects interrelate with regard to the study’s conclusions?
  o Are the ethical conclusions actually linked with normative premises? (e.g. avoiding an is/ought fallacy) [15,24]
  o Are the empirical conclusions supported by the data, or is there a bias in the empirical results based on the normative conclusions? (e.g. avoiding a normativist fallacy or "wishful thinking", deducing broad conclusions from fine-grained data, under- or overrating of empirical data, ignoring of empirical evidence that would criticize normative conclusions etc.) [84,85]
• Is there a critical evaluation of the results? (e.g. addressing methodological critique with regard to interdisciplinary cooperation, or indication of limitations)
Table 4: Criteria related to research ethics & scientific ethos

**Competing interests:**

- Which personal (e.g. cultural, philosophical, theological) presumptions concerning ethics may bias the EE research process significantly and how can they be adequately managed? (e.g. inclination to a emotivist meta-ethics, a neopositivist philosophy of science, a postmodernist account of society etc.) [93]

**Informed consent:**

- Do different standards exist in the various disciplines involved, and if so, have they been critically and respectfully discussed among the EE research team to find the most appropriate ethical standard? (e.g. is waiving of consent allowed, is assent sufficient, how to establish informed consent in emotional difficult situations at the end of life etc.) [91,94,95,96,97,99,100]
- Is the EE research team aware of a possible confidence bonus, and have strategies been developed to deal with this phenomenon carefully? (e.g. making transparent which goals and which limitations the own study will have and informing research participants and partners accordingly)

**Reporting results:**

- Is the EE research team aware of the (implicit) ethical impact of the way results are presented? (e.g. was the potential for stigmatization or discrimination considered when choosing the wording and emphasis of particular results?) [77, partly 87]
- Has the EE research team made sure that the way the results are presented reduces the potential for misinterpretation by third parties such as politicians and special interest groups? (e.g. by changing perspectives when re-reading results and revising the wording etc.)

**Consequences for the future:**

- Has the risk–benefit ratio for the EE research project been discussed in terms of its possible consequences for the people/society of the (near/more distant) future? (e.g. does lay considerable burden on study participants for a relatively low practical or epistemic output?) [77]
- Has the research team overlooked any negative consequences that could be detected in advance and therefore avoided? (e.g. acknowledging non-scientific partners when publishing, the handling of emotional distress of participants in interview studies with sensible questions, supervision of researchers involved in asking sensible questions etc.)
We believe we have now sufficiently addressed all concerns and critiques by the reviewers.
We would appreciate if the editors agree with our decisions for revision and approve finally the paper for publication in BMC Medical Ethics.

Sincerely, on behalf of all authors,
Yours Silke Schicktanz