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)	A comprehensive survey among statistical members of medical ethics committees in Germany on their personal impression of completeness and correctness of biostatistical aspects of submitted study protocols
AUTHORS	Rauch, Geraldine; Hafermann, Lorena; Mansmann, Ulrich; Pigeot, Iris

VERSION 1 - REVIEW

REVIEWER	Kevin Dobbin, Associate Professor of Biostatistics University of Georgia Athens, Georgia USA
REVIEW RETURNED	24-Jul-2019

GENERAL COMMENTS	The authors present the results of a survey of statistical members of medical ethics committees in Germany. My main issue is that why the readers should be interested in this survey could have been set up better by the authors. Why are these results important? For Item 2 in Table 1, it appears that multiple responses could be given by each participant. If so, this should be clarified. Also, the percentages presented result from dividing by 57, not the total number of responses, and this should be clarified. For Item 6 in Table 2, it appears that 10 out of 57 did not answer this key question. Some discussion of why this would happen seemed needed. There are a number of typographical errors and poorly constructed sentences. One problematic typo on page 12, first paragraph, second-to-last sentence, where it says "recalculate the sample size" it should read "reproduce the sample size"

REVIEWER	Allison Nugent National Institute of Mental Health United States of America
REVIEW RETURNED	28-Jul-2019

GENERAL COMMENTS	Overall, this manuscript would benefit greatly from careful editing by a native English speaker. There are numerous places throughout the manuscript where the intention of the authors is unclear. Abstract - Language throughout this section is unclear. - Questionnaire. "Urgency of further training" - training of whom? "assessment of completeness and correctness of these aspects" - completeness of these aspects in the protocol reviewed? - Primary and secondary outcome measures - this section seems to further describe the questionnaire, and does not state which measures were primary and which were secondary. -Results "the latter are in median affected 20-60% more often" - affected by what? - Conclusions - "freely formulated studies" - I'm not familiar with this terminology (although I am a US IRB member... Perhaps "studies not involving regulated drugs or devices?" It seems incorrect to call these "non-regulated", since they undergo ethics committee oversight. Introduction - Purpose or research question. It is unclear how this project "explores strategies towards a national standard for biostatistical reviews." - ".. standard of statistical reviews would even impose larger barriers" - barriers to what? - The authors suggest that they can identify topics that need "specific training for biostatisticians in medical ethics committees." Would additional training be required for the biostatistician review committee members, or the investigators submitting protocols? Methods -Data collection instruments and technologies - last paragraph - the intent of this last item is unclear. Is this question assessing the need for training of the statistician answering the survey? Or of investigators submitting surveys? -Data processing - what is meant by the statement that experts "validated" the survey? - The section on human subjects protections and confidentiality should be moved earlier in the methods section Results - On the first paragraph on Page 9, the authors talk about "need for refreshment." Does this refer to areas where investigators submitting protocols for review need additional training? - Table 2: I don't think that the translation of item 11 is accurate. I don't understand the last field for "additionally freely formulated aspects" - The authors state that studies under AMG/MPG regulatory requirement are of "higher statistical quality," although no actual statistical tests are performed to determine if these actually different in a statistically significant fashion. Given that the topic of this manuscript is the statistical quality of studies, this is a bit ironic. Discussion - Few consistent results from the literature are reported in the discussion. Given the degree to which biostatisticians noted problems with submitted studies, it would seem appropriate to

	discuss global difficulties with reproducibility across biomedical research.

REVIEWER	Jaranit Kaewkungwal Faculty of Tropical Medicine, Mahidol University, Thailand
REVIEW RETURNED	22-Aug-2019

GENERAL COMMENTS	The paper is well-written and interesting. The authors described the survey methodology and results clearly as well as recognized the limitations of the study. Although the study was about opinions of the statisticians in the ethics committees in Germany, it is informative and a lesson-learned for researchers elsewhere. One question that I have is about the question asking for the ratings on completeness and correctness of the 12 statistical issues by the statisticians, were all research proposals required to have all those 12 aspects? There might be many research proposals that did not need to have some of those 12 issues. In that case, if I was the respondent to this survey, I cannot estimate the x% correctly.

REVIEWER	Dario Gregori University of Padova, Italy
REVIEW RETURNED	30-Aug-2019

GENERAL COMMENTS	I'm quite enthusiastic about this work. This is a very relevant and well-conducted study. The authors present the results in a very appropriate way and they also discuss well the study limitations. My only suggestion is to add in the discussion some comparison with other analogous surveys in Europe. This could help the reader to put the results, now focused on Germany, in a broader context.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Reviewer Name: Kevin Dobbin, Associate Professor of Biostatistics

Institution and Country: University of Georgia, Athens, Georgia, USA

1. The authors present the results of a survey of statistical members of medical ethics committees in Germany. My main issue is that why the readers should be interested in this survey could have been set up better by the authors. Why are these results important?

Thank you for this important point. We integrated a paragraph on the motivation of our survey in the introduction section under “Purpose or research question”.

2. For Item 2 in Table 1, it appears that multiple responses could be given by each participant. If so, this should be clarified. Also, the percentages presented result from dividing by 57, not the total number of responses, and this should be clarified.

We added a footnote to clarify where multiple responses were possible.

3. For Item 6 in Table 2, it appears that 10 out of 57 did not answer this key question. Some discussion of why this would happen seemed needed.

Indeed, only 47 participants answered item 6, assessing if the statistical quality of ethical proposals generally differs between the regulated and the non-regulated setting. As this item was asked for in the survey before the specific biostatistical aspects were listed, this general question seemed to be difficult for a large part of the participants to answer. These considerations were added on page 10 of the revised manuscript.

4. There are a number of typographical errors and poorly constructed sentences.

Thank you for this important comment. The manuscript was revised by a native speaker and carefully corrected for typos.

5. One problematic typo on page 12, first paragraph, second-to-last sentence, where it says "recalculate the sample size" it should read "reproduce the sample size"

We corrected the sentence as proposed.

Reviewer 2

Reviewer Name: Allison Nugent

Institution and Country: National Institute of Mental Health, United States of America

• Overall, this manuscript would benefit greatly from careful editing by a native English speaker. There are numerous places throughout the manuscript where the intention of the authors is unclear.

Thank you for this important comment. The manuscript was now carefully revised by a native speaker.

- Abstract - Language throughout this section is unclear.

The wording of the abstract was carefully revised to achieve more clarity. It is difficult to find a good balance between clarity and keeping the very restricted word limit, but we did our best to solve this issue.

- Questionnaire. "Urgency of further training" - training of whom? "assessment of completeness and correctness of these aspects" - completeness of these aspects in the protocol reviewed?

- Primary and secondary outcome measures - this section seems to further describe the questionnaire, and does not state which measures were primary and which were secondary.

The corresponding paragraphs were extended to clarify the meaning.

- -Results "the latter are in median affected 20-60% more often" - affected by what?

- Conclusions - "freely formulated studies" - I'm not familiar with this terminology (although I am a US IRB member... Perhaps "studies not involving regulated drugs or devices?" It seems incorrect to call these "non-regulated", since they undergo ethics committee oversight.

Thank you for this comment. Indeed the expression “freely formulated” seem to be only familiar in German. We changed the wording throughout the revised manuscript. We now briefly introduce the term “non-regulated studies” as studies that are not regulated by law, i.e. by the German Medicines Act (AMG)/German Act on Medical Devices (MPG).

- Introduction

- Purpose or research question. It is unclear how this project "explores strategies towards a national standard for biostatistical reviews."

Thank you for pointing this out. We integrated a paragraph on the motivation of our survey in the introduction section under “Purpose or research question”.

- ".. standard of statistical reviews would even impose larger barriers" - barriers to what?

This sentence has been revised. Our focus was to argue that a national standard is required before an international standard can be achieved.

- The authors suggest that they can identify topics that need "specific training for biostatisticians in medical ethics committees." Would additional training be required for the biostatistician review committee members, or the investigators submitting protocols?

Indeed, this point was not made clear. Trainings can be addressed to both - to the statistical reviewers to improve clarity of statistical reviewer comments and to the medical investigators to improve their statistical knowledge. This has been revised in the text.

- Methods

- -Data collection instruments and technologies - last paragraph - the intent of this last item is unclear. Is this question assessing the need for training of the statistician answering the survey? Or of investigators submitting surveys?

Thank you for addressing this issue. The survey assessed the need for training of medical investigators. This has been revised in the text.

- -Data processing - what is meant by the statement that experts "validated" the survey?

- The section on human subjects protections and confidentiality should be moved earlier in the methods section

Thank you for this important note. This paragraph has been clarified and the section on human subjects protection and confidentiality has been moved.

- Results

- On the first paragraph on Page 9, the authors talk about "need for refreshment." Does this refer to areas where investigators submitting protocols for review need additional training?

- Table 2: I don't think that the translation of item 11 is accurate. I don't understand the last field for "additionally freely formulated aspects"

Indeed, the wording was not clear and we carefully reformulated these sentences.

- The authors state that studies under AMG/MPG regulatory requirement are of "higher statistical quality," although no actual statistical tests are performed to determine if these actually different in a statistically significant fashion. Given that the topic of this manuscript is the statistical quality of studies, this is a bit ironic.

In our mind, this is not ironic. This is an exploratory study (which is clearly stated on page 8) and not a hypothesis-driven confirmatory trial. The authors of this manuscript are all active as biostatisticians in medical research and from our experience it is one of the most prominent problems, that medical researchers tend to report lists of p-values even in an exploratory setting. The reported scale and location measures together with the box-plots are much more informative than a list of 24 p-values, where adjustment for multiplicity would not be meaningful. The reporting of p-values is a controversy issue, but we have a very strong position here. Therefore, we cannot follow this specific comment, in particular since there is an intensive discussion , , , going on about the abuse of p-values. We hope you agree with our decision.

• Discussion

- Few consistent results from the literature are reported in the discussion. Given the degree to which biostatisticians noted problems with submitted studies, it would seem appropriate to discuss global difficulties with reproducibility across biomedical research.

We agree with you that the discussion missed relevant result. We therefore added three important publications which we integrated in the discussion.

Reviewer 3

Reviewer Name: Jaranit Kaewkungwal

Institution and Country: Faculty of Tropical Medicine, Mahidol University, Thailand

The paper is well-written and interesting. The authors described the survey methodology and results clearly as well as recognized the limitations of the study. Although the study was about opinions of the statisticians in the ethics committees in Germany, it is informative and a lesson-learned for researchers elsewhere. One question that I have is about the question asking for the ratings on completeness and correctness of the 12 statistical issues by the statisticians, were all research proposals required to have all those 12 aspects? There might be many research proposals that did not need to have some of those 12 issues. In that case, if I was the respondent to this survey, I cannot estimate the x% correctly.

This is an important point and it's definitively true that there are studies where not all of the 12 aspects apply. We tried to address this issue in the "strengths and limitations" paragraph, were it says "This survey only classified study protocols as regulated by the AMG/MPG or non-regulated, where the latter covers a very heterogeneous range of studies for which the statistical requirements are not all the same." But we agree that this aspect should be repeated more prominently within the manuscript. We also clarified, that the participants had the option to choose "not assessable" for every individual aspect. Please see our changes on page 11 of the revised manuscript.

Reviewer 4

Reviewer Name: Dario Gregori

Institution and Country: University of Padova, Italy

[1] Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman, SN, Altman DG. Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016;31:337-350

[1] Ioannidis JPA. Publishing research with P-values: Prescribe more stringent statistical significance or proscribe statistical significance? Eur Heart J. 2019;40:2553-2554

[1] Ioannidis JPA. Options for publishing research without any P-values. Eur Heart J. 2019;40:2555-2556

[1] Szucs D, Ioannidis JPA. When null hypothesis significance testing is unsuitable for research: A reassessment. Front Hum Neurosci. 2017;11:390

I'm quite enthusiastic about this work. This is a very relevant and well-conducted study. The authors present the results in a very appropriate way and they also discuss well the study limitations.

My only suggestion is to add in the discussion some comparison with other analogous surveys in Europe. This could help the reader to put the results, now focused on Germany, in a broader context.

Thank you for this nice suggestion. We added additional references citing international surveys as proposed.

VERSION 2 – REVIEW

REVIEWER	Kevin Dobbin University of Georgia
REVIEW RETURNED	25-Oct-2019

GENERAL COMMENTS	The authors have addressed my concerns.

REVIEWER	Allison Nugent National Institute of Mental Health
REVIEW RETURNED	27-Nov-2019

GENERAL COMMENTS	The authors have sufficiently addressed all of my concerns and I have no further issues with publication.