BMJ Open

# 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



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2019-032864 |
| Article Type: | Research |
| Date Submitted by the Author: | 10-Jul-2019 |
| Complete List of Authors: | Rauch, Geraldine; Charité Universitätsmedizin Berlin, Institute of Biometry and Clinical Epidemiology<br>Hafermann, Lorena<br>Mansmann, Ulrich; Ludwig-Maximilians-Universitat Munchen, Departement of Medical Information Processing, Biometry, and Epidemiology<br>Pigeot, Iris; Leibniz Institute for Prevention Research and Epidemiology - BIPS, |
| Keywords: | MEDICAL ETHICS, STATISTICS & RESEARCH METHODS, EPIDEMIOLOGY |
| | |

**Title:** 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:** Geraldine Rauch[1,2*], Lorena Hafermann[1,2], Ulrich Mansmann[3], Iris Pigeot[4,5]

[1]Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute of Biometry and Clinical Epidemiology, Charitéplatz 1, 10117 Berlin, Germany.

[2]Berlin Institute of Health (BIH), Anna-Louisa-Karsch 2, 10178 Berlin, Germany

[3]Institute for Medical Information Processing, Biometry, and Epidemiology, Ludwig-Maximilian University Munich, Germany

[4]Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany

[5]University of Bremen, Institute of Statistics, Bremen, Germany

*Corresponding author's email: geraldine.rauch@charite.de

**Word count (excluding references):** 4560

## Abstract: (297 words)

**Objectives:** To assess completeness and correctness of study protocols submitted to medical ethics committees in Germany with respect to biostatistical aspects according to the personal appraisal of their statistical members.

**Design:** We conducted a web-based survey (31 items) among professional biostatisticians being active as members in German medical ethics committees during the past 3 years.

**Setting:** The study population was identified by a comprehensive web-search on all websites of German medical ethics committees.

**Participants:** The final contact list comprised 86 persons, who fulfilled the inclusion criterion. Of these 57 (66%) completed the survey.

**Questionnaire:** The first item checked the inclusion criterion, the last item assessed if the participant enjoyed the survey. Four items aimed at characterizing the specific medical ethics committee, one item asks for the urgency of further training on biostatistical aspects and 2x12 items asked for an individual assessment of the completeness and correctness of these aspects, while distinguishing studies according to the German Medicines Act (AMG)/German Act on Medical Devices (MPG) and non-regulated studies.

**Primary and secondary outcome measures:** To assess personal judgement, participants were asked to complete the sentence "In x% of the submitted study protocols, *the following problem occurs*", were 12 statistical issues were formulated (e.g. *specification of significance level is missing*).

**Results:** For all 12 biostatistical aspects, 45 of 49 (91.8%) participants judged the quality and completeness of AMG/MPG study protocols much better than that of non-regulated studies, where the latter are in median affected 20-60% more often. The highest need for training was reported for sample size calculation, missing values, and multiple comparison procedures.

**Conclusions:** Based on individual judgements of biostatisticians active in German medical ethics committees, the general completeness and correctness of biostatistical issues in study protocols is low for freely formulated studies and much better for AMG/MPG studies.

**Strengths and limitations of this study**

- This is the first survey among biostatisticians active in German medical ethics committees to assess the subjective judgement of quality and completeness regarding biostatistical aspects of study protocols.
- Although having put much effort in searching for all biostatisticians in German medical ethics committees, not all corresponding websites list their active and former members due to data protection reasons. Therefore, the target population was not completely identified.
- Confidentiality issues did not allow assessing the statistical quality of the content of individual study protocols submitted to German medical ethics committees. Therefore, only indirect information based on anonymized general personal judgement could be analyzed.
- This survey only classified study protocols as AMG/MPG or freely formulated studies (non-regulated). However, the latter class of studies covers a very broad range of studies including observational trials, retrospective analyses, surveys etc., for which the statistical requirements are not all the same. Therefore, the issues assessed by this survey might not be applicable to all types of studies.

- The review took place too early to study the impact on recent revisions on the statistical concepts of estimands.

**Funding statement.**

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. This paper was written on behalf of a recently established joint working group "Biometry in institutional review boards" of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS) e. V. and the German Region of the International Biometric Society. (https://gmds.de/aktivitaeten/medizinische-biometrie/arbeitsgruppenseiten/projektgruppen/biometrie-in-der-ethikkommission/, accessed July 2019)

**A competing interests statement**.

All authors declare no conflict of interest.

## Introduction

### *Problem formulation*

Medical Ethics Committees (Institutional review boards) aim to supervise the quality and validity of medical studies in order to ensure an ethically justifiable positive benefit-risk profile i.[1-3] Members of medical ethics committees do not only assess the submitted material but also act as consultants. Besides the medical content aspects, the medical ethics committee also looks at the legal and scientific validity including biostatistical aspects of the study design and analysis strategy.[4-6] Although general guidance for a good biostatistical practice in medical research projects does exist, there is no consensus and only limited guidance to what extent medical ethics committees should assess these statistical aspects.[5-8] According to the last revision of the German Drug Regulation Law in 2016 (Bundesgesetzblatt (2016), § 41a), a biostatistician is a mandatory member for a medical ethics committees, next to medical as well as legal experts, and lay persons.[9] However, not all medical ethics committees are involved in studies regulated by law and therefore not required to appoint a biostatistician.[10-13] Moreover, medical research is faced with new challenges related to the digitalization of the health system and the focus on personalized medicine, which also brings along new tasks and perspectives for medical ethics committees.[14]

### *Purpose or research question*

This project aimed to explore strategies towards a national standard for biostatistical reviews of study protocols within a medical ethics committee. As medical ethics committee organization and composition already vary in Germany, an international standard of statistical reviews would even impose larger barriers. Therefore, we conducted a comprehensive survey among biostatisticians active in German medical ethics committees between 2016 and 2018. The aim is to evaluate and quantify the personal judgement of the participants on the quality and completeness of statistical aspects in clinical study protocols submitted to German medical ethics committees.

A direct judgement of the statistical quality of study protocols would have required the assessment of relevant protocol extracts or even entire study protocols by the experts. This design failed caused by enforced data protection mechanisms. Many medical ethics committees argued that original protocols (even partly and anonymized) could not be made available for the planned assessment without impairing the trust which is an essential part of the medical ethics committee's standing.

To overcome this problem, we decided to ask biostatisticians in medical ethics committees to give a global, personal assessment on specific issues of the statistical quality and completeness of study protocols. On the one hand, the individual impression does not objectively reflect the "true" quality. On the other hand, objective quality criteria are very hard to define and would definitively impose a need for a controversial discussion. Therefore, the individual global quality assessment of biostatisticians in medical ethics committees provides an informative marker to at least roughly assess current standards and problems. Biostatisticians in medical ethics committees review many study protocols and can well reflect the statistical problems currently met. From these findings, we can identify statistical topics which need an enforced focus, e.g. within the framework of GCP courses, specific training for biostatisticians in medical ethics committees, or in future guidance papers.[15-17]

## Methods

### *Qualitative approach and research paradigm*

This study is a comprehensive systematic survey among biostatisticians who were members of German medical ethics committees between 2016 and 2018.

### *Researcher characteristics and reflexivity*

The questionnaire was developed by a senior biostatistician, reviewed and extended by five independent biostatisticians including two professors of biostatistics, two senior biostatisticians, and one bachelor student with a limited background in biostatistics. The latter person was consulted in particular to assess the comprehensibility of the wording.

### *Context*

All persons involved in developing the survey, analyzing the data, and writing this article are members of the joint project group "Biometry in ethics committees" of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS) e. V. and the German Region of the International Biometric Society (https://gmds.de/aktivitaeten/medizinische-biometrie/arbeitsgruppenseiten/projektgruppen/biometrie-in-der-ethikkommission/, accessed June 2019). This group was founded in 2017 and aims to strengthen the work of biostatisticians in medical ethics committees by offering specific training (in methods as well as communication of statistical issues to non-statisticians), establishing a communication network for mutual support, and developing specific guidelines which allow a standardized, high quality statistical review of study protocols.

### *Data collection instruments and technologies*

The questionnaire was implemented as an online survey (www.umfrageonline.com, accessed Juli 2019). The survey consisted of 31 items, which were grouped in 11 steps/pages in the online survey. Questions were formulated in German. The original survey can be found here (https://www.umfrageonline.com/s/6b2e8f4&preview=1&DO-NOT-SEND-THIS-LINK-ITS-ONLY-PREVIEW, accessed July 2019). The original questions in English translation are provided in Appendix 1.

The first item checked the central inclusion criteria if the respondent served as statistical expert in a medical ethics committee within the last three years. Only persons who answered this question positively were included in the final analysis. The last item evaluated if the respondent enjoyed the survey. Of the remaining 29 main items, four items characterized specific features of the medical ethics committee and the review process within the medical ethics committee. We asked (1) for the medical ethics committee type (ethics committee of a medical faculty, of a State Chamber of Physicians (Landesärztekammer) or other), (2) for the federal state in Germany, where the medical ethics committee is located, (3) how many studies per year the respondent reviews on average (in steps of 50) and (4) if the respondent is exclusively responsible for study proposals according to the German Medicines Act (AMG)/German Act on Medical Devices (MPG) or also for freely formulated, non-regulated studies. In case the respondent´s medical ethics committee is responsible for regulated as well as for non-regulated studies, another (conditional) item asked whether the statistical quality of study protocols is better in the regulated compared to the non-regulated setting.

Additionally, 2 times 12 items asked for assessing the completeness and correctness of different biostatistical aspects (12 for the regulated setting, 12 for the non-regulated setting conditional on the responsibilities of the specific medical ethics committee as marked in the previous item). Participants were asked to complete the sentence "In x% of the submitted study protocols, *the following problem occurs*", were 12 statistical topics were listed (e.g. *specification of the significance level is missing*). Participants could provide percentages in steps of 10% (0% to 100%) with a higher percentage indicating a worse result. In principle, items formulated in the way "In x% of the submitted study protocols, *the following problem occurs*" could have also been assessed on an interval scale by allowing for continuous specifications of percentages. This would have pretended a quantitative and objective assessment. However, the subjective impression is surely neither quantitative nor completely objective.

An additional item asked for the need to refresh the statistical knowledge of the applicant on a certain topic to be selected out of a list of nine statistical topics with the option to add additional ones. In addition, this need had to be assessed as low, medium or high.

***Data processing***

The online survey system saved the answers of the participants in a central database where data can be downloaded in various formats. An extended group of experts including the authors validated the online survey and final corrections were integrated.

***Units of study***

The study population is defined as all biostatisticians being members of a German medical ethics committee between 2016 and 2018.

***Sampling strategy***

To identify the study population, a web search on the homepages of all German medical ethics committees was performed which resulted in a preliminary email list. All these email addresses were freely available on the web. Moreover, several persons known to be active in ethics committees were asked to complete this list in agreement with the specific biostatistician. The final list of eligible candidates consisted of 86 biostatisticians.

***Data collection methods***

The call to participate in the survey was sent out by email on November 28, 2018 with a reminder on December 13, 2018. The survey was opened until December 15, 2018. Some participants actively asked for the possibility to slightly extend the deadline on which we agreed. The original survey was, however, open only between December 1, 2018 and January 31, 2019, so that after that time period data collection was completed.

***Techniques to enhance trustworthiness***

The webpage for the survey was not publically published to avoid participation of persons not belonging to the study population. Still, in principle, anyone who was aware of the link could have participated in the survey. There is no way to check for fulfillment of the inclusion criterion or correctness of the provided answers. However, as the link was not easily available and the study population was likely to be highly compliant, this risk seems to be minor. The survey could be completed only once from a single IP address. In principle, participants could have completed the survey several times using different IP addresses. As the survey was anonymous due to data

protection reasons, it is impossible to verify such frauds. However, as there was no benefit in completing the survey more than once, this risk seems to be minor, too. We did not advertise the survey by means of global mailing lists within biostatistical societies, although this approach was discussed. The reason for preferring a limited and focused mailing list is that otherwise no response proportion could have been evaluated which is, however, a crucial quality indicator of a survey. Moreover, a global mailing list would have included a large proportion of recipients who did not fulfil the eligibility criteria.

***Ethical issues pertaining to human subjects***

The online survey was anonymized and most items included an option for providing no answer to this specific question. Some items, like the question asking for the medical ethics committee's federal state, allowed for a potential re-identification of the respondent in case of a small federal state like Bremen with perhaps only one medical ethics committee and further knowledge on the medical ethics committee's members, but respondents were free to leave this box blank. No personal data were collected from the participants. Participation in the survey was voluntary and not rewarded. To enable reproduction of the results presented in this paper, the final dataset is freely available from the Dryad repository (doi:10.5061/dryad.5v3v47d) without any information on the medical ethics committee's federal state to avoid any risk of potential re-identification. For this voluntary survey in healthy participants without any risk of putting harm to the respondents and without any direct medical research focus, an ethical approval was not necessary.

***Data analysis***

This is an exploratory study, which is analyzed using descriptive statistical methods. Items 1 to 6 as well as items 30 and 31 are simple categorical items assessed on a multiple choice basis. The items were evaluated by means of absolute and relative frequencies. The 2x12 items asking for the assessment of the completeness and correctness of different biostatistical aspects are likert-scaled ordinal variables with 11 possible outcomes (0%, 10%, …, 100%). For these items, we reported absolute and relative frequencies and graphically displayed them as stacked bar charts. Moreover, we provided medians, quartiles and grouped boxplots, where two groups of studies are considered, namely those conducted in a regulatory setting and those who are non-regulated (freely formulated). All analyses were performed using the statistical software R, version 3.5.1. The original dataset (excluding information on the German federal state in which the specific medical ethics committee is located) is freely available from the Dryad repository (doi:10.5061/dryad.5v3v47d) to allow for reproducibility.

***Patient and Public Involvement***

This survey does not include patients or the general public. The design and the development of the survey was intensively discussed by the members of the joint project group "Biometry in ethics committees" of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS) e. V. and the German Region of the International Biometric Society (https://gmds.de/aktivitaeten/medizinische-biometrie/arbeitsgruppenseiten/projektgruppen/biometrie-in-der-ethikkommission/, accessed July 2019).

## Results

Table 1 shows the characteristics of the medical ethics committees to which the 57 participants of the survey are appointed. Note that the number of answers differ per item, as the online survey

offered the option to abstain from answering a specific question. The majority of participants 46 (80.7%) were members of an ethics committees of a medical faculty of a university, whereas 15 (26.3%) were members of an ethics committee of a State Chamber of Physicians (Landesärztekammer). Some participants were members of more than one medical ethics committee at the same time.

A total of 47 participants answered the question on the location of the medical ethics committee to which they were appointed as member. The medical ethics committees are located in 12 out of 16 German federal states where 14 (28.6%) participants were members of medical ethics committees in Northrhine-Westphalia and 7 (14.3%) in Baden-Württemberg.

A total of 18 (32.1%) participants reviewed up to 50 study proposals per year, 14 participants (25.0%) reviewed between 51 and 100 study proposals per year and 24 participants (42.0%) reviewed more than 100 study proposals on average per year.

The vast majority of participants 50 (89%) reviewed both – study proposals according to AMG/MPG and freely formulated study proposals.

| **Item 2: In which type of ethics committee have you been active?** | **N=57 (%)** |
|---|---|
| medical ethics committee of a medical faculty of a university | 46 (80.7) |
| medical ethics committee of a State Chamber of Physicians (Landesärztekammer) | 15 (26.3) |
| Others | 7 (12.3) |
| **Item 3: Which German federal state is the medical ethics committees located to which you were affiliated?** | **N=50 (%)** |
| Baden-Württemberg | 7 (14.3) |
| Bavaria | 4 (8.2) |
| Berlin | 3 (6.1) |
| Brandenburg | 0 (0.0) |
| Bremen | 0 (0.0) |
| Hamburg | 1 (2.0) |
| Hesse | 3 (6.1) |
| Mecklenburg Western Pomerania | 0 (0.0) |
| Lower Saxony | 6 (12.2) |
| Northrhine-Westphalia | 14 (28.6) |
| Rhineland Palatinate | 2 (4.1) |
| Saarland | 0 (0.0) |
| Saxony | 1 (2.0) |
| Saxony-Anhalt | 4 (8.2) |
| Schleswig-Holstein | 3 (6.1) |
| Thuringia | 2 (2.0) |
| **Item 4: How many study proposals (amendments and annual reports excluded) do you review on average per year?** | **N=56 (%)** |
| Up to 50 | 18 (32.1) |
| 51-100 | 14 (25.0) |
| 101-150 | 7 (12.5) |
| 151-200 | 7 (12.5) |
| >200 | 10 (17.9) |
| **Item 5: Which kinds of studies do you review as member of a medical ethics committee?** | **N=56 (%)** |
| Studies referring to AMG or MPG | 3 (5.4) |
| Freely formulated studies | 3 (5.4) |

| | |
|---|---|
| Both | 50 (89.3) |

**Table 1 Characteristics of the medical ethics committees to which the participants are appointed.**

With respect to the general biostatistical quality of study protocols, Table 2 displays the results of Items 6 and 11. Out of 47 participants in total who responded to Item 6, a majority of 45 (95.7%) stated that study protocols under regulatory requirements (AMG/MPG) are on average of higher statistical quality compared to studies without such requirements. The remaining 2 (4.3%) participants stated that there is no difference on average. Item 11 asked how the participants considered the need for refreshing specific biostatistical topics (see Table 2) according to the applications. A high need for a refreshment was especially identified for "Handling of missing values" as indicated by 34 participants (75.6%), for "Sample size calculation" as indicated by 27 participants (60.0%), for "Multiple comparison procedures" 26 (57.8%) and for "Adjustment for covariables" 26 (57.8%). For all topics, at least 70% of participants considered the need for a refreshment of statistical knowledge as middle or high.

| **Item 6: Do you in general have the impression that the biometrical quality of ethical proposals for studies referring to AMG/MPG compared to proposals for studies not referring to AMG/MPG differs?** | **N=47 (%)** |
|---|---|
| Yes, ethical proposals for studies referring to AMG/MPG have a higher biometrical quality on average | 45 (91.8) |
| Yes, ethical proposals for studies not referring to AMG/MPG have a higher biometrical quality on average | 0 (0.0) |
| No, the biometrical quality does not differ on average | 2 (4.1) |
| **Item 11: How do you consider the need for refreshment of the following biostatistical topics for applicants at medical ethics committees?** | **N=45 (%)** |
| **Study design** | |
| Low | 7 (15.6) |
| Middle | 23 (51.1) |
| High | 15 (33.3) |
| **Wording of the study aims, hypotheses and/or endpoints** | |
| Low | 1 (2.2) |
| Middle | 22 (48.9) |
| High | 22 (48.9) |
| **Sample size calculation** | |
| Low | 2 (4.4) |
| Middle | 15 (33.3) |
| High | 27 (60.00) |
| Not assessable | 1 (2.2) |
| **Differentiation between confirmatory and exploratory analyses** | |
| Low | 4 (8.9) |
| Middle | 17 (37.8) |
| High | 24 (53.3) |
| **Handling of missing values** | |
| Low | 2 (4.4) |
| Middle | 9 (20.00) |
| High | 34 (75.6) |
| **Description of the statistical analysis** | |
| Low | 1 (2.3) |
| Middle | 23 (52.3) |

| | |
|---|---|
| High | 20 (45.5) |
| **Multiple comparisons problems** | |
| Low | 3 (6.7) |
| Middle | 16 (35.6) |
| High | 26 (57.8) |
| **Adjustment for covariables** | |
| Low | 3 (6.7) |
| Middle | 18 (40.0) |
| High | 24 (53.3) |
| **Randomization/stratification** | |
| Low | 10 (22.2) |
| Middle | 27 (60.00) |
| High | 6 (13.3) |
| Not assessable | 2 (4.4) |
| **Additional freely formulated aspects** | |
| Data management | |
| Regulatory view | |
| Testing versus estimating | |

**Table 2. General biostatistical quality of study proposals.**

Items 7 to 10 asked to assess completeness and correctness of biostatistical aspects while distinguishing studies according to the regulatory setting (Arzneimittelgesetz, AMG, Medizinproduktegesetz, MPG) (Items 7 and 8) and studies without regulatory requirements (Items 9 and 10). Participants were asked to complete the sentence "*In x% of the submitted study protocols, the following problem occurs*", where 12 statistical issues were formulated (e.g. *specification of the significance level is missing*). Participants could give the percentage in steps of 10% (0% to 100%) with a higher percentage indicating a worse result. The results referring to these items are presented in Table 3 and displayed in Figure 1 as grouped boxplots. Additionally, Figure 2 displays the percentages of the item categories for both study types as staggered bar plots.

| **In x[%] of the ethical proposals, *** | **Median (1.; 3. quartile)** | |
|---|---|---|
| | **Regulated studies (AMG, MPG)** | **Non-regulated studies (freely formulated)** |
| 1)… wording of study aims, hypotheses and/or endpoints are inadequate or inconsistent. | 10 (10;20) | 50 (20; 80) |
| 2)… specification of the significance level is missing. | 10 (0;10) | 30 (10; 50) |
| 3)… sample size calculation is incomplete, inadequate or missing completely. | 20 (10;30) | 70 (50; 80) |
| 4)… only general statements on statistical analysis methods are provided not fitting and addressing the specific study aim. | 10 (10;25) | 70 (50; 80) |
| 5)… no clear differentiation between confirmatory and explanatory analyses is provided. | 15 (10; 27.5) | 70 (50; 82.5) |
| 6)… specification of one- or two-sided statistical testing is missing. | 10 (10; 30) | 55 (30; 87.5) |
| 7)… description of how to handle missing values is incomplete, inadequate or missing completely. | 40 (22.5; 70) | 90 (80;90) |
| 8)… description of how to handle multiple comparisons problems is incomplete, inadequate or | 30 (20; 50) | 80 (70;90) |

| | | |
|---|---|---|
| missing completely. | | |
| 9)… description of how and for which covariables adjustment is planned is incomplete, inadequate or missing completely. | 25 (20; 50) | 80 (50;90) |
| 10)… specification of randomization and stratification is incomplete, inadequate or missing completely. | 10 (10; 30) | 65 (50;80) |
| 11)… no study biometrician is specified. | 20 (10; 50) | 70 (40;90) |
| 12)… description of statistical methods is not sufficiently specified. | 20 (10; 40) | 80 (60;90) |

*Number of participants varied per item and study type between N=47 and N=51.

**Table 3. Medians and quartiles of percentages of assessments for completeness and correctness of statistical aspects for AMG studies in comparison to non-regulated, freely formulated studies (higher values indicate less completeness and/or less correctness).**

It becomes obvious that freely formulated studies tend to be of much lower statistical quality and show less completeness, regardless of the specific topic. Differences in medians of the ratings between studies with regulatory (yes/no) requirements range between 20% and 60%. The statistical aspects "missing values", "multiple testing problems" as well as "adjustment for covariables" show the highest discrepancies between both study types. For instance, the statistical methods are not sufficiently specified in 80% on average (median) for non-regulated study proposals whereas this is the case in only 20% on average (median) for studies with regulatory requirements. Similarly, for non-regulated studies only general statements on statistical analysis methods are provided not fitting and addressing the specific study aim in 70% on average (median) whereas this is only stated in 10% on average (median) for regulated studies. For non-regulated studies, all 12 statistical aspects are indicating as showing high deficiencies, while only in 10% on average (median) of all proposals of regulated studies aspects are mentioned which have not been completely or correctly addressed in the proposals.

*Please include here Figure 1*

*Please include here Figure 2*

## Discussion

This systematic survey among biostatisticians serving in German medical ethics committees aimed to assess the individual impression of completeness and correctness of biostatistical aspects of submitted study protocols. As an overall result, the completeness and correctness of handling statistical issues in the submitted study protocols is heterogeneous. There is a notably difference in quality between study protocols with and without regulatory requirements, where the latter show major deficits. A specifically high need for refreshment was especially identified for "Handling of missing values", "Sample size calculation", "Multiple comparison procedures", and "Adjustment for covariables". However, for non-regulated studies there also exist quite general deficiencies, as the description of the statistical methods is not sufficiently specified or only general statements on statistical analysis methods are provided not fitting and addressing the specific study aim. It should be mentioned that for regulatory studies the ICH E9 guideline offers guidance how to analyze a clinical study.[6] This guideline is also helpful for the non-regulated setting but may not be known to members of this community.[13]

To the best of our knowledge, this is the first survey among biostatisticians being a member in German medical ethics committees and the first attempt to measure the quality and completeness of biostatistical issues in medical study protocols. Wang et al. (2018) conducted a survey among consulting biostatisticians with respect to the quality of reporting the statistical analysis strategy after data were already analyzed.[18] The four most frequently reported statistical problems in reporting were "removing or altering some data records to better support the research hypothesis", "interpreting the statistical findings on the basis of expectation, not actual results", "not reporting the presence of key missing data that might bias the results", and "ignoring violations of assumptions that would change results from positive to negative". Clark et al. (2013) screened original study protocols submitted to UK ethics committees for the completeness and correctness of sample size derivation.[19] They found that only 42% of the study protocols reported all information which is required to accurately recalculate the sample size. These findings are in line with the results of our survey.

A total of 56 (66%) of the contacted 86 persons participated and fulfilled the inclusion criteria of having been an active member as biostatistician in a medical ethics committee in the past three years. This corresponds to a high participation proportion. Though, we do not know if we identified all relevant people.

A further limitation of our survey is that it does not provide an overview of the true, objectively measured quality and completeness of biostatistical aspects of study protocols but that it only refers to the subjective, individual impression of the statistical members of the ethics committees. An objective rating of the study protocols, however, was not possible due to data and privacy protection issues, as this would have required screening of the submitted study documents. Moreover, objective measures for quality are difficult to define as the views on "adequate" quality can differ considerably. The subjective ratings of completeness and correctness are subject to inter- and intra-rater variability. Therefore, the results should rather be interpreted as a rough indication and not as definite numbers.

As a third limitation we consider the fact that the survey could not assess recent issues added in an addendum to the ICH E9 guideline.[20] It presents a structured framework to link trial objectives to a suitable trial design and tools for estimation and hypothesis testing. This framework introduces the concept of an estimand, translating the trial objective into a precise definition of the treatment effect that is to be estimated. It also aims to facilitate the dialogue between disciplines involved in a clinical trial.

Even in view of these limitations, the survey clearly indicates the need for basic and advanced statistical trainings and guidance for medical researchers. All medical faculties in Germany have established biostatistical units providing consulting services. However, this does not seem to be sufficient to enable medical researches to develop protocols, which cover statistical issues adequately. Reasons could be that medical researchers undervalue the impact of an appropriate biometrical planning on the quality and validity of medical studies. In personal discussions with participants of this survey, several of them reported frequent examples where the statistical analysis strategy in study protocols is addressed with a single general sentence like "The data are analyzed with valid statistical methods." This does not only indicate a lack of statistical knowledge but also a lack of awareness that statistical methods have an important impact on the validity of medical research.

The survey also gauges the range of methodological challenges encountered by biostatisticians being member of a medical ethics committee. Often a biostatistician is the last methodological sentinel before a study is implemented in a clinical setting. In order to involve more biostatistician in medical ethics committees, there is a need to provide support to enable them to adequately discharge their responsibilities. Unfortunately, the survey did not check how the remaining members of the medical ethics committee react on revision requests by biostatisticians and if these requests are adequately addressed before the final vote of the medical ethics committee on the criticized study protocol. Moreover, we did not formally assess if biostatisticians who are members of medical ethics committees formulate their requests in comparable detail and persistence. Due to own experiences and based on narrative reports we suspect that biostatistical concerns cannot be easily communicated to and understood by the non-statistical members of the medical ethics committee. Thus, there is a need to establish a better communication, which allows expressing biostatistical concerns in a convincing lay language?

It is therefore time to communicate the general importance of statistics for medical research. This includes the establishment of guidelines for protocol writing and templates like the SPIRIT Statement, which also handles statistical input to a protocol.[21,22] Furthermore, the implementation of reporting guidelines like STROBE should be made more popular.[23] Moreover, the development of specific trainings and guidance on how to address specific statistical challenges is required. Finally, national standards for the tasks of a biostatistician as a member of a medical ethics committee must be formulated.[24,25]

## References


1. General Assembly of the World Medical Association. (2014). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *The Journal of the American College of Dentists* 81: 14.
2. Doppelfeld E, Hasford J (2019): Medizinische Ethikkommissionen in der Bundesrepublik Deutschland: Entstehung und Einbindung in die medizinische Forschung. *Bundesgesundheitsbla* 62: 682-689.
3. Buchner B, Hase F, Borchers D, Pigeot I (2019): Aufgaben, Regularien und Arbeitsweise von Ethikkommissionen. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, *62*(6), 690-696.
4. Gelfond JA, Heitman E, Pollock BH, Klugman CM (2011): Principles for the ethical analysis of clinical and translational research. *Stat Med* 30:2785-2792.
5. Schlattmann P, Scherag A, Rauch G, Mansmann U (2019): Biometrische Aspekte in Ethikkommissionen. *Bundesgesundheitsbla* 62: 751-757.
6. International Conference on Harmonization (1999) ICH E9 Guideline. Statistical principles for clinical trials. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf (accessed 09. Juli 2019)
7. Raspe H, Hüppe A, Strech D, J T (2012): Empfehlungen zur Begutachtung klinischer Studien durch Ethik-Kommissionen. Deutscher Ärzte-Verlag.
8. Victor N (1999): Prüfung der wissenschaftlichen Qualität und biometriespezifischer Anforderungen durch die Ethikkommission. *MedR Medizinrecht* 17:408-412.
9. Bundesgesetzblatt (2016): Viertes Gesetz zur Änderungen arzneimittelrechtlicher und anderer Vorschriften, 20. Dezember 2016, Teil I Nr.63

10. Gesetz über den Verkehr mit Arzneimitteln (Arzneimittelgesetz−AMG). https://www.gesetze-im-internet.de/mpg/MPG.pdf (accessed 09. Juli 2019)
11. Gesetz über Medizinprodukte (Medizinproduktegesetzz−MPG). https://www.pei.de/SharedDocs/Downloads/gesetze/arzneimittelgesetz.pdf?__blob=publicationFile&v=6. (accessed 09. Juli 2019)
12. Felden K (1994): Nationale und Europäische Richtlinien für klinische Prüfungen. In: AH HJS (ed) Arzneimittelprüfung und Good Clinical Practice. MMV-Verag, München.
13. Pigeot I, Foraita R, Frömke C, Haller B, Hoffmann VS, Zapf A (2019): Ethische Bewertung von Studien am Menschen außerhalb des regulatorischen Rahmens: nicht bindend, aber von großer Wichtigkeit. *Bundesgesundheitsbla* 62: 722-728.
14. Rauch G, Röhmel J, Gerß J, Scherag A, Hofner B (2019): Aktuelle Herausforderungen bei der Bewertung von Ethikanträgen–Aspekte der Digitalisierung und Personalisierung im Gesundheitswesen. *Bundesgesundheitsbla* 62: 758-764.
15. Good P, Hardin J (2009) Common errors in statistics (and how to avoid them). Wiley
16. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwarz Lm, Woloshin S (2008): Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest* 8:53-96.
17. Ioannidis JP, Greenland S, Hlatky MA, et al. (2014): Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet* 383:166-175.
18. Wang MQ, Yan AF, Katz RV (2018) Researcher Rrequests for nappropriate Aanalysis and reporting: A U.S. survey of consulting biostatisticians. *Ann Intern Med* 169:554-558.
19. Clark T, Berger U, Mansmann U (2013): Sample size determinations in original research protocols for randomised clinical trials submitted to UK research ethics committees: review. BMJ 346:f1135.
20. International Conference on Harmonization (2017) ICH E9 (R1) Addendum. ICH E9 (R1) Addendum on estimands and sensitivity5 analysis in clinical trials to the guideline on statistical 6 principles for clinical trials https://www.ema.europa.eu/en/documents/scientific-guideline/draft-ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical_en.pdf (accessed 09. Juli 2019)
21. Chan AW, Tetzlaff JM, Altman DG, Laupacis A et al. (2013): SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med* 158: 200-207.
22. Clark T, Davies H, Mansmann U (2014): Five questions that need answering when considering the design of clinical trials. *Trials* 15: 286.
23. Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 147: 573-577.
24. Zapf A, Huebner M, Rauch G, Kieser M (2018): What makes a biostatistician? *Stat Med* 38:695-701
25. Gerlinger C, Edler L, Friede T et al. (2012): Considerations on what constitutes a 'qualified statistician' in regulatory guidelines. *Stat Med* 31:1303-1305.

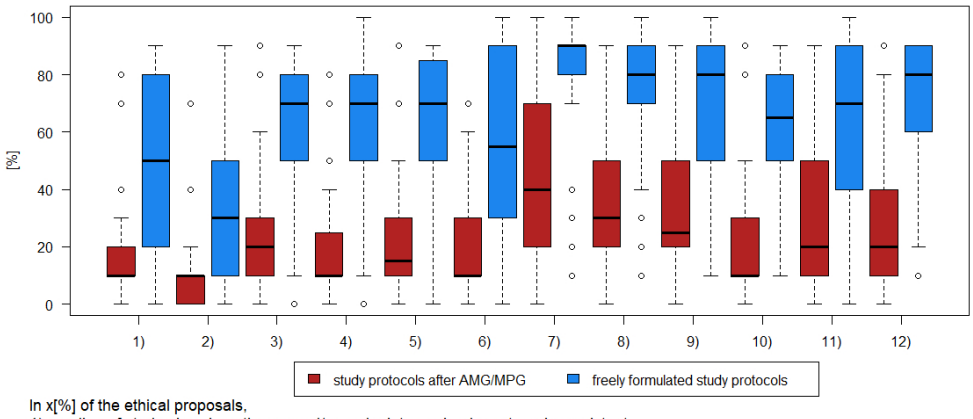
**Author's contribution**

- Geraldine Rauch developed the research idea, designed the survey and wrote the manuscript
- Lorena Hafermann implemented the online survey, performed the analyses and reviewed the manuscript
- Ulrich Mansmann helped to develop the research idea, reviewed the survey, reviewed the manuscript

- Iris Pigeot helped to develop the research idea, reviewed the survey, reviewed the manuscript

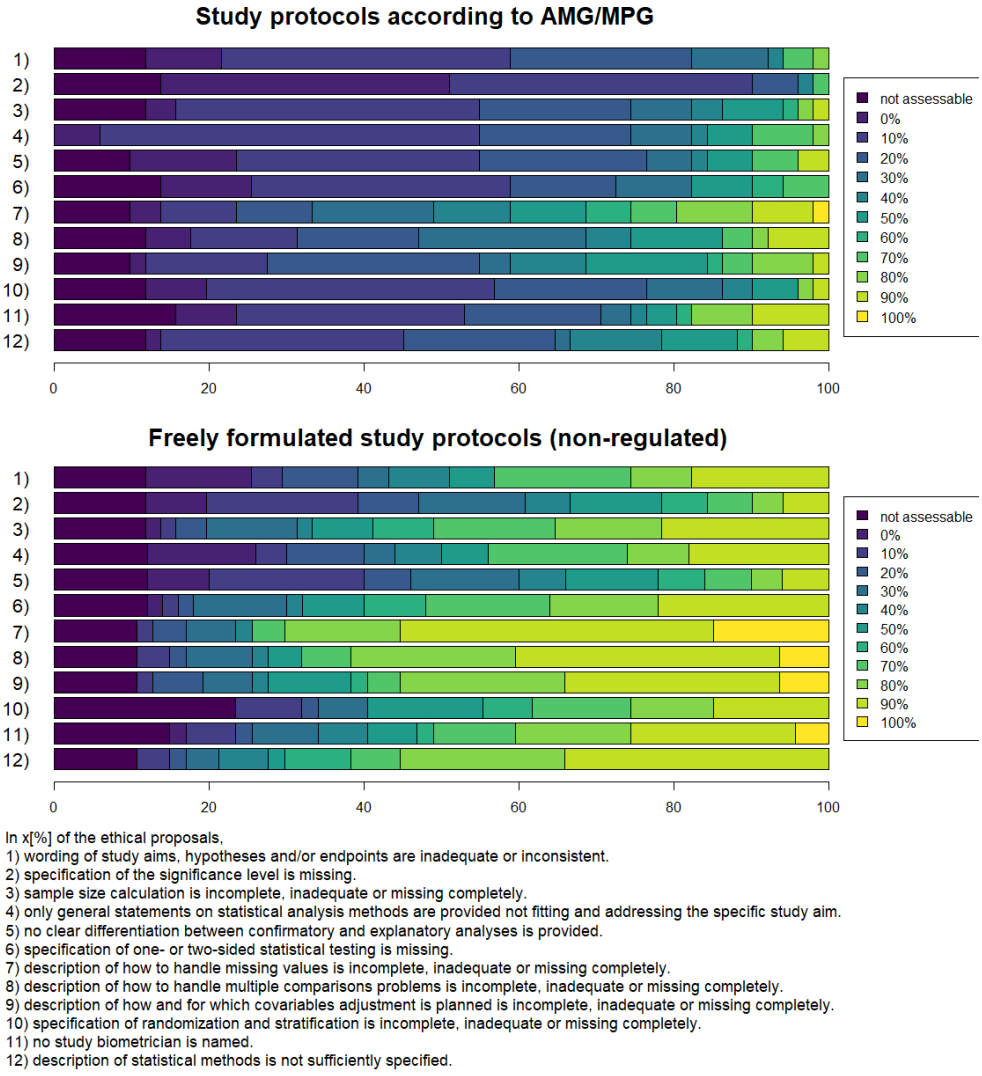
**A data sharing statement**

The original dataset of the survey is available from the Dryad repository, doi:10.5061/dryad.5v3v47d.




In x[%] of the ethical proposals,
1) wording of study aims, hypotheses and/or endpoints are inadequate or inconsistent.
2) specification of the significance level is missing.
3) sample size calculation is incomplete, inadequate or missing completely.
4) only general statements on statistical analysis methods are provided not fitting and addressing the specific study aim.
5) no clear differentiation between confirmatory and explanatory analyses is provided.
6) specification of one- or two-sided statistical testing is missing.
7) description of how to handle missing values is incomplete, inadequate or missing completely.
8) description of how to handle multiple comparisons problems is incomplete, inadequate or missing completely.
9) description of how and for which covariables adjustment is planned is incomplete, inadequate or missing completely.
10) specification of randomization and stratification is incomplete, inadequate or missing completely.
11) no study biometrician is named.
12) description of statistical methods is not sufficiently specified.

Boxplots of completeness and correctness of statistical aspects for AMG studies (red) and non-regulated, freely formulated studies (blue).




Completeness and correctness of statistical aspects in percentages for AMG/MPG studies (upper part) and non-regulated, freely formulated studies (lower part).

# Original questions of the survey in in English translation

**1) Do you currently work or have you been working as biometrician/statistician in a medical ethics commission?**

- yes
- no

**2) In which kinds of medical ethic committees have you been working?**

- medical ethic committee of a medical faculty of a university
- medical ethic committee of a State Chamber of Physicians (Landesärztekammer)
- others

**3) Which German federal state is the medical ethics committees located to which you were affiliated?**

- Baden-Württemberg
- Bavaria
- Berlin
- Brandenburg
- Bremen
- Hamburg
- Hesse
- Mecklenburg Western Pomerania
- Lower Saxony
- Northrhine - Westphalia
- Rhineland Palatinate
- Saarland
- Saxony - Anhalt
- Schleswig Holstein
- Thuringia
- Saxony

**4) How many study proposals (amendments and annual reports excluded) do you review on average per year?**

- up to 50
- 51 - 100
- 101 - 150
- 151 - 200
- more than 200

**5) Which kinds of studies do you review as member of a medical ethics committee?**

- Studies referring to AMG (German Drug Act) or MPG (German Medical Products Act)
- Other studies not referring to AMG/MPG

**6) Do you in general have the impression that the biometrical quality of ethical proposals for studies referring to AMG/MPG compared to proposals for studies not referring to AMG/MPG differs?**

- Yes, ethical proposals for studies referring to AMG/MPG have a higher biometrical quality on average
- Yes, ethical proposals for studies not referring to AMG/MPG have a higher biometrical quality on average
- No, the biometrical quality does not differ on average
- Not specified

**7) Please complete the following statements to biometrical points of criticism in ethical proposals for studies referring to AMG/MPG through specification of approximate percentages according to your subjective assessment.**

**Part 1**

- In x[%] of the ethical proposals, wording of study aims, hypotheses and/or endpoints are inadequate or inconsistent.
- In x[%] of the ethical proposals, specification of the significance level is missing.
- In x[%] of the ethical proposals, sample size calculation is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals, only general statements on statistical analysis methods are provided not fitting and addressing the specific study aim.
- In x[%] of the ethical proposals, no clear differentiation between confirmatory and explanatory analyses is provided.
- In x[%] of the ethical proposals, specification of one- or two-sided statistical testing is missing.

**8) Part 2**

- In x[%] of the ethical proposals, description of how to handle missing values is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals, description of how to handle multiple comparisons problems is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals, description of how and for which covariables adjustment is planned is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals, specification of randomization and stratification is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals no study biometrician is named.
- In x[%] of the ethical proposals, description of statistical methods is not sufficiently specified.

**9) Please complete the following statements to biometrical points of criticism in ethical proposals for studies not referring to AMG/MPG through specification of approximate percentages according to your subjective assessment.**

**Part 1**

- In x[%] of the ethical proposals, wording of study aims, hypotheses and/or endpoints are inadequate or inconsistent.
- In x[%] of the ethical proposals, specification of the significance level is missing.
- In x[%] of the ethical proposals, sample size calculation is incomplete, inadequate or missing completely.

- In x[%] of the ethical proposals, only general statements on statistical analysis methods are provided not fitting and addressing the specific study aim.
- In x[%] of the ethical proposals, no clear differentiation between confirmatory and explanatory analyses is provided.
- In x[%] of the ethical proposals, specification of one- or two-sided statistical testing is missing.

**10) Part 2**

- In x[%] of the ethical proposals, description of how to handle multiple comparisons problems is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals, description of how and for which covariables adjustment is planned is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals, specification of randomization and stratification is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals no study biometrician is named.
- In x[%] of the ethical proposals, description of statistical methods is not sufficiently specified.

**11) How do you consider the need for refreshment of the following biostatistical topics for applicants at medical ethics committees? (low, middle, high, not assessable)**

- Study design
- Wording of the study aims, hypotheses and/or endpoints
- Sample size calculation
- Differentiation between confirmatory and exploratory analyses
- Handling of missing values
- Description of the statistical analysis
- Multiple comparisons problems
- Adjustment for covariables
- Randomization/stratification

**12) Did you like our survey?**

- yes
- no

# Checklist "Standards for Reporting Qualitative Research (SRQR)"*

This manuscript describes a quantitative survey. None of the provided checklists perfectly fits this type of empirical study. We felt that the checklist "Standards for Reporting Qualitative Research (SRQR)*" in many aspects addresses our type of study, although this is not a qualitative survey.

## Standards for Reporting Qualitative Research (SRQR)*

http://www.equator-network.org/reporting-guidelines/srqr/

| | Page/line no(s). |
|---|---|
| **Title and abstract** | |
| **Title** - Concise description of the nature and topic of the study Identifying the study as qualitative or indicating the approach (e.g., ethnography, grounded theory) or data collection methods (e.g., interview, focus group) is recommended | Page 1 |
| **Abstract** - Summary of key elements of the study using the abstract format of the intended publication; typically includes background, purpose, methods, results, and conclusions | Page 2 |
| **Introduction** | |
| **Problem formulation** - Description and significance of the problem/phenomenon studied; review of relevant theory and empirical work; problem statement | Page 4 |
| **Purpose or research question** - Purpose of the study and specific objectives or questions | Page 4 |
| **Methods** | |
| **Qualitative approach and research paradigm** - Qualitative approach (e.g., ethnography, grounded theory, case study, phenomenology, narrative research) and guiding theory if appropriate; identifying the research paradigm (e.g., postpositivist, constructivist/ interpretivist) is also recommended; rationale** | Page 5 |
| **Researcher characteristics and reflexivity** - Researchers' characteristics that may influence the research, including personal attributes, qualifications/experience, relationship with participants, assumptions, and/or presuppositions; potential or actual interaction between researchers' characteristics and the research questions, approach, methods, results, and/or transferability | Page 5 |
| **Context** - Setting/site and salient contextual factors; rationale** | Page 5 |
| **Sampling strategy** - How and why research participants, documents, or events were selected; criteria for deciding when no further sampling was necessary (e.g., sampling saturation); rationale** | Page 5 |
| **Ethical issues pertaining to human subjects** - Documentation of approval by an appropriate ethics review board and participant consent, or explanation for lack thereof; other confidentiality and data security issues | Page 5 |

| | |
|---|---|
| **Data collection methods** - Types of data collected; details of data collection procedures including (as appropriate) start and stop dates of data collection and analysis, iterative process, triangulation of sources/methods, and modification of procedures in response to evolving study findings; rationale** | Page 5 |
| **Data collection instruments and technologies** - Description of instruments (e.g., interview guides, questionnaires) and devices (e.g., audio recorders) used for data collection; if/how the instrument(s) changed over the course of the study | Page 5 |
| **Units of study** - Number and relevant characteristics of participants, documents, or events included in the study; level of participation (could be reported in results) | Page 6 |
| **Data processing** - Methods for processing data prior to and during analysis, including transcription, data entry, data management and security, verification of data integrity, data coding, and anonymization/de-identification of excerpts | Page 6 |
| **Data analysis** - Process by which inferences, themes, etc., were identified and developed, including the researchers involved in data analysis; usually references a specific paradigm or approach; rationale** | Page 7 |
| **Techniques to enhance trustworthiness** - Techniques to enhance trustworthiness and credibility of data analysis (e.g., member checking, audit trail, triangulation); rationale** | Page 6 |

**Results/findings**

| | |
|---|---|
| **Synthesis and interpretation** - Main findings (e.g., interpretations, inferences, and themes); might include development of a theory or model, or integration with prior research or theory | Pages 7-11 |
| **Links to empirical data** - Evidence (e.g., quotes, field notes, text excerpts, photographs) to substantiate analytic findings | x |

**Discussion**

| | |
|---|---|
| **Integration with prior work, implications, transferability, and contribution(s) to the field -** Short summary of main findings; explanation of how findings and conclusions connect to, support, elaborate on, or challenge conclusions of earlier scholarship; discussion of scope of application/generalizability; identification of unique contribution(s) to scholarship in a discipline or field | Pages 11-12 |
| **Limitations** - Trustworthiness and limitations of findings | Pages 12-13 |

**Other**

| | |
|---|---|
| **Conflicts of interest** - Potential sources of influence or perceived influence on study conduct and conclusions; how these were managed | Page 3 |
| **Funding** - Sources of funding and other support; role of funders in data collection, interpretation, and reporting | Page 3 |

*The authors created the SRQR by searching the literature to identify guidelines, reporting standards, and critical appraisal criteria for qualitative research; reviewing the reference lists of retrieved sources; and contacting experts to gain feedback. The SRQR aims to improve the transparency of all aspects of qualitative research by providing clear standards for reporting qualitative research.

**The rationale should briefly discuss the justification for choosing that theory, approach, method, or technique rather than other options available, the assumptions and limitations implicit in those choices, and how those choices influence study conclusions and transferability. As appropriate, the rationale for several items might be discussed together.

BMJ Open

# 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



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2019-032864.R1 |
| Article Type: | Original research |
| Date Submitted by the Author: | 11-Oct-2019 |
| Complete List of Authors: | Rauch, Geraldine; Charité Universitätsmedizin Berlin, Institute of Biometry and Clinical Epidemiology; Berlin Institute of Health<br>Hafermann, Lorena; Charité Universitätsmedizin Berlin, Institute of Biometry and Clinical Epidemiology; Berlin Institute of Health<br>Mansmann, Ulrich; Ludwig-Maximilians-Universitat Munchen, Departement of Medical Information Processing, Biometry, and Epidemiology<br>Pigeot, Iris; Leibniz Institute for Prevention Research and Epidemiology - BIPS, |
| Primary Subject Heading: | Research methods |
| Secondary Subject Heading: | Ethics, Epidemiology |
| Keywords: | MEDICAL ETHICS, STATISTICS & RESEARCH METHODS, EPIDEMIOLOGY |
| | |

**Title:** 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:** Geraldine Rauch[1,2*], Lorena Hafermann[1,2], Ulrich Mansmann[3], Iris Pigeot[4,5]

[1]Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute of Biometry and Clinical Epidemiology, Charitéplatz 1, 10117 Berlin, Germany.

[2]Berlin Institute of Health (BIH), Anna-Louisa-Karch 2, 10178 Berlin, Germany

[3]Institute for Medical Information Processing, Biometry, and Epidemiology, Ludwig-Maximilian University Munich, Germany

[4]Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany

[5]University of Bremen, Institute of Statistics, Bremen, Germany

*Corresponding author's email: geraldine.rauch@charite.de

**Word count (excluding references):** 4972

## Abstract: (299 words)

**Objectives:** To assess biostatistical quality of study protocols submitted to German medical ethics committees according to personal appraisal of their statistical members.

**Design:** We conducted a web-based survey among biostatisticians who have been active as members in German medical ethics committees during the past 3 years.

**Setting:** The study population was identified by a comprehensive web-search on websites of German medical ethics committees.

**Participants:** The final list comprised 86 eligible persons. In total, 57 (66%) completed the survey.

**Questionnaire:** The first item checked whether the inclusion criterion was met. The last item assessed satisfaction with the survey. Four items aimed to characterize the medical ethics committee in terms of type and location, one item asked for the urgency of biostatistical training addressed to the medical investigators. The main 2x12 items reported an individual assessment of the quality of biostatistical aspects in the submitted study protocols, while distinguishing studies according to the German Medicines Act (AMG)/German Act on Medical Devices (MPG) and studies non-regulated by these laws.

**Primary and secondary outcome measures:** The individual assessment of the quality of biostatistical aspects corresponds to the primary objective. Thus, participants were asked to complete the sentence "In x% of the submitted study protocols, *the following problem occurs*", where 12 different statistical problems were formulated. All other items assess secondary endpoints.

**Results:** For all biostatistical aspects, 45 of 49 (91.8%) participants judged the quality of AMG/MPG study protocols much better than that of "non-regulated" studies. The latter are in median affected 20-60% more often by statistical problems. The highest need for training was reported for sample size calculation, missing values, and multiple comparison procedures.

**Conclusions:** Biostatisticians being active in German medical ethics committees classify the biostatistical quality of study protocols as low for "non-regulated" studies, whereas quality is much better for AMG/MPG studies.

**Strengths and limitations of this study**

- This is the first survey among biostatisticians active in German medical ethics committees to assess the individual assessment the quality of and the completeness of information on biostatistical aspects in the submitted study protocols.
- Although having put much effort in searching for all biostatisticians active in German medical ethics committees, the target population was not completely identified.
- Confidentiality issues did not allow direct and objective assessment of individual study protocols' content.
- This survey classified study protocols as regulated by the AMG/MPG and those non-regulated by these laws, where the latter covers a very heterogeneous group of studies for which the statistical requirements are not all the same.
- This survey was conducted too early to study the impact on recent revisions on the statistical concepts of estimands.

**Funding statement**

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. This paper was written on behalf of a recently established joint working group "Biometry in institutional review boards" of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS) e. V. and the German Region of the International Biometric Society. (https://gmds.de/aktivitaeten/medizinische-biometrie/arbeitsgruppenseiten/projektgruppen/biometrie-in-der-ethikkommission/, accessed September 2019)

**A competing interests statement**

All authors declare no conflict of interest.

## Introduction

### *Problem formulation*

Medical Ethics Committees (institutional review boards) aim to judge the quality and validity of medical studies in order to ensure an ethically justifiable positive benefit-risk profile.[1-3] Their members do not only assess the submitted material but also act as consultants. Besides medical content, the board verifies legal and scientific validity including biostatistical aspects of the study design and analysis strategy.[4-6] Although general guidance for a good biostatistical practice in medical research projects exist, there is no consensus and only limited guidance to what extent medical ethics committees should assess these statistical aspects.[5-8] According to the last revision of the German Drug Regulation Law in 2016 (Bundesgesetzblatt (2016), § 41a), a biostatistician is a mandatory member of a medical ethics committee, beside of medical as well as legal experts, and lay persons.[9] However, not all medical ethics committees appraise legally regulated studies in which case a biostatistician is not mandatory.[10-13] Moreover, medical research is faced with new challenges related to the digitalization of the health system and the focus on personalized medicine, which also brings along new tasks and perspectives for medical ethics committees.[14]

### *Purpose or research question*

To increase the biostatistical quality of study protocols, standards for biostatistical reporting and for biostatistical reviewer comments have to be implemented in Germany that account for the fact that the organization and composition of German medical ethics committee organization are quite heterogeneous. Of course, on the long-run international standard have to be agreed upon. To achieve this global aim, the first step is to assess the current level of statistical quality of submitted study proposals, so that gaps and challenges can be identified.

For this purpose, we conducted a comprehensive survey among biostatisticians who were active in German medical ethics committees between 2016 and 2018. The aim is to evaluate and quantify the personal assessment of the participants of the quality and completeness of statistical aspects in clinical study protocols submitted to German medical ethics committees.

A direct judgement of the statistical quality of study protocols would have required the assessment of relevant protocol extracts or even entire study protocols by the experts. This was, however, not possible due to enforced data protection mechanisms. Many medical ethics committees argued that original protocols (even partly and anonymized) could not be made available for the planned assessment without impairing the trust which is an essential part of the medical ethics committee's standing.

To overcome this problem, we decided to ask biostatisticians in medical ethics committees to give a global, personal assessment on specific issues of the statistical quality and completeness of study protocols. On the one hand, the individual impression does not objectively reflect the "true" quality. On the other hand, objective quality criteria are very hard to define and would definitively impose a need for a controversial discussion. Therefore, the individual global quality assessment of biostatisticians in medical ethics committees provides an informative marker to at least roughly assess current standards and problems. Biostatisticians in medical ethics committees review many study protocols and can well reflect the statistical problems currently met. From these findings, we can identify statistical topics which need an enforced focus, e.g. within the framework of GCP (Good Clinical Practice) courses, specific training addressed to the statistical reviewers to improve clarity of statistical reviewer comments and other training addressed to the medical investigators to improve their statistical knowledge.[15-17]

## Methods

### *Qualitative approach and research paradigm*

This study is a comprehensive systematic survey among biostatisticians who were members of German medical ethics committees between 2016 and 2018.

### *Researcher characteristics and reflexivity*

The questionnaire was developed by a senior biostatistician, reviewed and extended by five independent biostatisticians including two professors of biostatistics, two senior biostatisticians, and one bachelor student with a limited background in biostatistics. The latter person was consulted in particular to assess the comprehensibility of the wording.

### *Context*

All authors who developed the survey, analyzed the data, and wrote this article are members of the joint project group "Biometry in ethics committees" of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS) e. V. and the German Region of the International Biometric Society (https://gmds.de/aktivitaeten/medizinische-biometrie/arbeitsgruppenseiten/projektgruppen/biometrie-in-der-ethikkommission/, accessed September 2019). This group was founded in 2017 and aims to strengthen the work of biostatisticians in medical ethics committees by offering specific training (in methods as well as communication of statistical issues to non-statisticians), establishing a communication network for mutual support, and developing specific guidelines which allow a standardized, high quality statistical review of study protocols.

### *Data collection instruments and technologies*

The questionnaire was implemented as an online survey (www.umfrageonline.com, accessed September 2019). The survey consisted of 31 items, which were grouped in 11 steps/pages in the online survey. Questions were formulated in German. The original survey can be found here (https://www.umfrageonline.com/s/6b2e8f4&preview=1&DO-NOT-SEND-THIS-LINK-ITS-ONLY-PREVIEW, accessed September 2019). English translations are provided in Appendix 1.

The first item checked the key inclusion criterion if the respondent served as statistical expert in a medical ethics committee within the last three years. Only persons who answered this question positively were included in the final analysis. The last item evaluated if the respondent enjoyed the survey. Of the remaining 29 main items, four items characterized specific features of the medical ethics committee and the review process within the medical ethics committee. We asked (1) for the type of the medical ethics committee (ethics committee of a medical faculty, of a State Chamber of Physicians (Landesärztekammer) or other), (2) for the federal state in Germany, where the medical ethics committee is located, (3) how many studies the respondent reviews on average per year (in steps of 50) and (4) if the respondent is exclusively responsible for study proposals according to the German Medicines Act (AMG)/German Act on Medical Devices (MPG) or also for studies that are non-regulated by these laws which will be referred to in the following briefly as non-regulated studies. In case the respondent´s medical ethics committee is responsible for regulated as well as for non-regulated studies, another (conditional) item asked whether the statistical quality of study protocols is better in the regulated compared to the non-regulated setting.

Additionally, 2x 12 items asked for assessing the completeness and correctness of different biostatistical aspects (12 for the regulated setting, 12 for the non-regulated setting conditional on the responsibilities of the specific medical ethics committee as marked in the previous item). Participants were asked to complete the sentence "In x% of the submitted study protocols, *the following problem occurs*", were 12 statistical problems were addressed (e.g. *"Specification of the significance level is missing"*). Participants could provide percentages in steps of 10% (0% to 100%) with a higher percentage indicating a worse result. In principle, items formulated in the way "In x% of the submitted study protocols, *the following problem occurs*" could have also been assessed on an interval scale by allowing for continuous specifications of percentages. This would have pretended a quantitative and objective assessment. However, the subjective impression is surely neither quantitative nor completely objective.

An additional item asked for the need to refresh the statistical knowledge of the, e.g., medical or epidemiological investigator on a certain topic to be selected out of a list of nine statistical topics with the option to add additional ones. In addition, this need had to be assessed as low, medium or high.

***Data processing***

The online survey system saved the answers of the participants in a central database where data can be downloaded in various formats. An extended group of experts including the authors validated the online survey by testing and commenting it. Final corrections were integrated after the validation phase.

***Units of study***

The study population is defined as all biostatisticians being members of a German medical ethics committee between 2016 and 2018.

***Sampling strategy***

To identify the study population, a web search on the homepages of all German medical ethics committees was performed which resulted in a preliminary email list. All these email addresses were freely available on the web. Moreover, several persons known to be active in ethics committees were asked to complete this list in agreement with the specific biostatistician. The final list of eligible candidates consisted of 86 biostatisticians.

***Data collection methods***

The call to participate in the survey was sent out by email on November 28, 2018 with a reminder on December 13, 2018. The survey was opened until December 15, 2018. Some participants actively asked for the possibility to slightly extend the deadline on which we agreed. The original survey was, however, open only between December 1, 2018 and January 31, 2019, so that after that time period data collection was completed.

***Techniques to enhance trustworthiness***

The webpage for the survey was not publically published to avoid participation of persons not belonging to the study population. Still, in principle, anyone who was aware of the link could have participated in the survey. There is no way to check for fulfillment of the inclusion criterion or correctness of the provided answers. However, as the link was not easily available and the study population was likely to be highly compliant, this risk seems to be minor. The survey could be

completed only once from a single IP address. In principle, participants could have completed the survey several times using different IP addresses, which seems very unlikely. As the survey was anonymous due to data protection reasons, it is impossible to verify such fraud. However, as there was no benefit in completing the survey more than once, this risk seems to be minor. We did not advertise the survey by means of global mailing lists within biostatistical societies, although this approach was discussed. A limited and focused mailing list is preferable as otherwise no responders' proportion could have been evaluated which is a crucial quality indicator of a survey. Moreover, a global mailing list would have included a large proportion of recipients who did not fulfil the eligibility criteria.

***Ethical issues pertaining to human subjects***

The online survey was anonymized and most items included an option for providing no answer. The question asking for the medical ethics committee's federal state allowed for a potential re-identification of the respondent in case of a small federal state like Bremen. Respondents were therefore free to leave this box blank. No personal data were collected from the participants. Participation in the survey was voluntary and not rewarded. To enable reproduction of the results presented in this paper, the final dataset is freely available from the Dryad repository (https://datadryad.org/stash/share/VjofDJkoUtqIjjaJQ84O7FB1W5cJu78g04cNey044no), without any information on the medical ethics committee's federal state to avoid any risk of potential re-identification. No ethical approval was necessary for this voluntary survey in healthy participants without any risk of putting harm to the respondents and without any direct medical research focus.

***Data analysis***

This is an exploratory study, which is analyzed using descriptive statistical methods. Items 1 to 6 as well as items 30 and 31 are simple categorical items assessed on a multiple-choice basis. The items were evaluated by means of absolute and relative frequencies. The 2x12 items asking for the assessment of the completeness and correctness of different biostatistical aspects are Likert-scaled ordinal variables with 11 possible outcomes (0%, 10%, …, 100%). For these items, we reported absolute and relative frequencies and graphically displayed them as stacked bar charts. Moreover, we provided medians, quartiles and grouped boxplots, where two groups of studies are considered (regulated versus non-regulated studies). All analyses were performed using the statistical software R, version 3.5.1. The original dataset (excluding information on the specific German federal state) is freely available from the Dryad repository (https://datadryad.org/stash/share/VjofDJkoUtqIjjaJQ84O7FB1W5cJu78g04cNey044no) to allow for reproducibility.

***Patient and public involvement***

This survey does not include patients or the general public. The design and the development of the survey was intensively discussed by the members of the joint project group "Biometry in ethics committees" of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS) e. V. and the German Region of the International Biometric Society.

## Results

Table 1 shows the characteristics of the medical ethics committees to which the 57 participants of the survey are appointed. Note that the number of answers differ per item, as the online survey offered the option to abstain from answering a specific question. A majority of 46 participants (80.7%) were members of an ethics committees of a medical faculty of a university, whereas 15

(26.3%) were members of an ethics committee of a State Chamber of Physicians (Landesärztekammer). Some participants were members of more than one medical ethics committee at the same time.

A total of 47 participants answered the question on the location of the medical ethics committee to which they were appointed as member. The medical ethics committees are located in 12 out of 16 German federal states, where 14 (28.6%) participants were members of medical ethics committees in Northrhine-Westphalia and 7 (14.3%) in Baden-Württemberg.

A total of 18 (32.1%) participants reviewed up to 50 study proposals per year, 14 participants (25.0%) reviewed between 51 and 100 study proposals per year and 24 participants (42.0%) reviewed more than 100 study proposals on average per year.

The vast majority of 50 participants (89%) reviewed both – study proposals according to AMG/MPG and study proposals in a non-regulated setting.

| **Item 2: In which type of ethics committee have you been active?*** | **N=57 (%)** |
|---|---|
| Medical ethics committee of a medical faculty of a university | 46 (80.7) |
| Medical ethics committee of a State Chamber of Physicians (Landesärztekammer) | 15 (26.3) |
| Others | 7 (12.3) |
| **Item 3: Which German federal state is the medical ethics committees located to which you were affiliated?*** | **N=50 (%)** |
| Baden-Württemberg | 7 (14.3) |
| Bavaria | 4 (8.2) |
| Berlin | 3 (6.1) |
| Brandenburg | 0 (0.0) |
| Bremen | 0 (0.0) |
| Hamburg | 1 (2.0) |
| Hesse | 3 (6.1) |
| Mecklenburg Western Pomerania | 0 (0.0) |
| Lower Saxony | 6 (12.2) |
| Northrhine-Westphalia | 14 (28.6) |
| Rhineland Palatinate | 2 (4.1) |
| Saarland | 0 (0.0) |
| Saxony | 1 (2.0) |
| Saxony-Anhalt | 4 (8.2) |
| Schleswig-Holstein | 3 (6.1) |
| Thuringia | 2 (2.0) |
| **Item 4: How many study proposals (amendments and annual reports excluded) do you review on average per year?** | **N=56 (%)** |
| Up to 50 | 18 (32.1) |
| 51-100 | 14 (25.0) |
| 101-150 | 7 (12.5) |
| 151-200 | 7 (12.5) |
| >200 | 10 (17.9) |
| **Item 5: Which kinds of studies do you review as member of a medical ethics committee?*** | **N=56 (%)** |
| Studies regulated by AMG or MPG | 3 (5.4) |
| Studies in a non-regulated setting | 3 (5.4) |
| Both | 50 (89.3) |

***Multiple responses possible (in case of membership in several ethics committees)**

**Table 1 Characteristics of the medical ethics committees to which the participants are appointed.**

With respect to the general biostatistical quality of study protocols, Table 2 displays the results of Items 6 and 11. Only 47 participants answered Item 6, assessing if the statistical quality of ethical proposals generally differs between regulated and the non-regulated studies. As this item was placed in the survey before the specific biostatistical aspects were named, this general formulation seemed to be difficult to understand for a large part of the participants. Out of 47 participants in total who responded to Item 6, a majority of 45 (95.7%) stated that study protocols under regulatory requirements (AMG/MPG) are on average of higher statistical quality compared to studies without such requirements. The remaining 2 (4.3%) participants stated that there is no difference on average. Item 11 asked how the participants considered the need for additional training in different statistical areas addressed to the investigators submitting protocols (see Table 2). A high need for a training was especially identified for *"Handling of missing values"* as indicated by 34 participants (75.6%), for *"Sample size calculation"* as indicated by 27 participants (60.0%), for *"Multiple comparison procedures"* 26 (57.8%) and for *"Adjustment for covariables"* 26 (57.8%). For all topics, at least 70% of participants considered the need for a refreshment of statistical knowledge as middle or high.

| **Item 6: Do you in general have the impression that the biostatistical quality of ethical proposals for studies regulated by AMG/MPG differs compared to proposals in a non-regulated setting?** | **N=47 (%)** |
|---|---|
| Yes, ethical proposals for studies regulated by AMG/MPG have a higher biostatistical quality on average | 45 (91.8) |
| Yes, ethical proposals for studies in a non-regulated setting have a higher biostatistical quality on average | 0 (0.0) |
| No, the biostatistical quality does not differ on average | 2 (4.1) |
| **Item 11: How do you consider the need for additional training in the following biostatistical areas for investigators submitting protocols to medical ethics committees?** | **N=45 (%)** |
| **Study design** | |
| Low | 7 (15.6) |
| Middle | 23 (51.1) |
| High | 15 (33.3) |
| **Wording of the study aims, hypotheses and/or endpoints** | |
| Low | 1 (2.2) |
| Middle | 22 (48.9) |
| High | 22 (48.9) |
| **Sample size calculation** | |
| Low | 2 (4.4) |
| Middle | 15 (33.3) |
| High | 27 (60.00) |
| Not assessable | 1 (2.2) |
| **Differentiation between confirmatory and exploratory analyses** | |
| Low | 4 (8.9) |
| Middle | 17 (37.8) |
| High | 24 (53.3) |
| **Handling of missing values** | |
| Low | 2 (4.4) |
| Middle | 9 (20.00) |

| | |
|---|---|
| High | 34 (75.6) |
| **Description of the statistical analysis** | |
| Low | 1 (2.3) |
| Middle | 23 (52.3) |
| High | 20 (45.5) |
| **Multiple comparisons problems** | |
| Low | 3 (6.7) |
| Middle | 16 (35.6) |
| High | 26 (57.8) |
| **Adjustment for covariables** | |
| Low | 3 (6.7) |
| Middle | 18 (40.0) |
| High | 24 (53.3) |
| **Randomization/stratification** | |
| Low | 10 (22.2) |
| Middle | 27 (60.00) |
| High | 6 (13.3) |
| Not assessable | 2 (4.4) |
| **Additional aspects (added as free-text response by the participants)** | |
| Data management | |
| Regulatory view | |
| Testing versus estimating | |

**Table 2. General biostatistical quality of study proposals.**

Items 7 to 10 asked to assess completeness and correctness of biostatistical aspects while distinguishing studies according to the regulatory setting (Arzneimittelgesetz, AMG, Medizinproduktegesetz, MPG) (Items 7 and 8) and studies without regulatory requirements (Items 9 and 10). Participants were only able to judge those study types (regulated and/or non-regulated) that were specified in item 5. Participants were asked to complete the sentence "*In x% of the submitted study protocols, the following problem occurs*", where 12 statistical problems were formulated (e.g. "*Specification of the significance level is missing"*). Participants could give the percentage in steps of 10% (0% to 100%) with a higher percentage indicating a worse result. Note that, study protocols submitted to German ethics committees can cover various types of research including observational studies, retrospective analyses, surveys, etc. Not all of the 12 aspects formulated below are applicable to all types of studies. The requested percentages refer to the average of all studies, which have been reviewed by the participant. However, there also was the option to classify a specific aspect as "not assessable". As a consequence, the number of valid responses vary per item, where lower values might indicate items that are more difficult to judge. Moreover, some participants interrupted the survey after the assessment of some items, probably because the statistical problems are repeated for regulated and non-regulated studies, which might have decreased the motivation. The results referring to the valid responses are presented in Table 3 and displayed in Figure 1 as grouped boxplots. Additionally, Figure 2 displays the percentages of the item categories for both study types as staggered bar plots.

| **In x[%] of the ethical proposals,** | **Median (1.; 3. quartile), n= Number of valid responses (excluding "non assessable")** | |
|---|---|---|
| | **Regulated studies (AMG, MPG)** | **Non-regulated studies** |
| 1)… wording of study aims, hypotheses and/or endpoints are inadequate or inconsistent. | 10 (10;20), n=45 | 50 (20; 80), n=45 |
| 2)… specification of the significance level is missing. | 10 (0;10), n=44 | 30 (10; 50) , n=45 |

| | | |
|---|---|---|
| 3)… sample size calculation is incomplete, inadequate or missing completely. | 20 (10;30), n=45 | 70 (50; 80) , n=45 |
| 4)… only general statements on statistical analysis methods are provided not fitting and addressing the specific study aim. | 10 (10;25), n=46 | 70 (50; 80), n=45 |
| 5)… no clear differentiation between confirmatory and explanatory analyses is provided. | 15 (10; 27.5), n=46 | 70 (50; 82.5), n=45 |
| 6)… specification of one- or two-sided statistical testing is missing. | 10 (10; 30), n=44 | 55 (30; 87.5), n=43 |
| 7)… description of how to handle missing values is incomplete, inadequate or missing completely. | 40 (22.5; 70), n=46 | 90 (80;90), n=42 |
| 8)… description of how to handle multiple comparisons problems is incomplete, inadequate or missing completely. | 30 (20; 50), n=45 | 80 (70;90), n=42 |
| 9)… description of how and for which covariables adjustment is planned is incomplete, inadequate or missing completely. | 25 (20; 50), n=46 | 80 (50;90), n=42 |
| 10)… specification of randomization and stratification is incomplete, inadequate or missing completely. | 10 (10; 30), n=45 | 65 (50;80), n=36 |
| 11)… no study biometrician is specified. | 20 (10; 50), n=43 | 70 (40;90), n=40 |
| 12)… description of statistical methods is not sufficiently specified. | 20 (10; 40), n=45 | 80 (60;90), n=42 |

**Table 3. Medians and quartiles of percentages of assessments for completeness and correctness of statistical aspects for AMG studies in comparison to non-regulated studies (higher values indicate a lower level of completeness and/or correctness).**

It turns out that protocols of non-regulated studies tend to be of much lower statistical quality and show a lower level of completeness, regardless of the specific topic. Differences in medians of the ratings between regulated and non-regulated studies range between 20% and 60%. The statistical aspects *"missing values"*, *"multiple testing problems"* as well as *"adjustment for covariables"* show the highest discrepancies between both study types. For instance, the statistical methods are not sufficiently specified in 80% on average (median) for non-regulated study proposals whereas this is the case in only 20% on average (median) for studies with regulatory requirements. Similarly, for non-regulated studies only general statements on statistical analysis methods are provided not fitting and addressing the specific study aim in 70% on average (median), whereas this is only stated in 10% on average (median) for regulated studies. For non-regulated studies, all 12 statistical aspects show high deficiencies, while only in 10% on average (median) of all proposals of regulated studies aspects are mentioned which have not been completely or correctly addressed in the proposals.

*Please include here Figure 1*

*Please include here Figure 2*

## Discussion

This systematic survey among biostatisticians serving in German medical ethics committees aimed to assess the individual impression of completeness and correctness of biostatistical aspects of submitted study protocols. As an overall result, the completeness and correctness of handling statistical issues in the submitted study protocols is heterogeneous. There is a notably difference in quality between study protocols with and without regulatory requirements, where the latter show

major deficits. A specifically high need for refreshment was identified for *"Handling of missing values"*, *"Sample size calculation"*, *"Multiple comparison procedures"*, and *"Adjustment for covariables"*. However, there also exist quite general deficiencies for non-regulated studies, as the description of the statistical methods is not sufficiently specified. It should be mentioned that for regulatory studies the ICH E9 guideline offers guidance on how to analyze a clinical study.[6] This guideline is also helpful for the non-regulated settings but may be unknown to members of this community.[13]

To the best of our knowledge, this is the first survey among biostatisticians who were members of German medical ethics committees and the first attempt to assess the quality and completeness of biostatistical issues in medical study protocols. Wang et al. (2018) conducted a survey among biostatistical consultants with respect to the quality of reporting the statistical analysis strategy after data was already analyzed.[18] The four most frequently reported statistical problems were *"Removing or altering some data records to better support the research hypothesis"*, *"Interpreting the statistical findings on the basis of expectation, not actual results"*, *"Not reporting the presence of key missing data that might bias the results"*, and *"Ignoring violations of assumptions that would change results from positive to negative"*. Clark et al. (2013) screened original study protocols submitted to UK ethics committees for the completeness and correctness of sample size derivation.[19] They found that only 42% of the study protocols reported all information, which is required to accurately reproduce the sample size. Kilkenny et al. (2009) conducted a survey of the quality of experimental design, statistical analysis and reporting of research in animal studies.[20] They found that only 59% of the studies stated the hypothesis or objective of the study and the number and characteristics of the animals used und only 70% of the publications described their methods and presented the results with a measure of error or variability.[20] These findings are in line with the results of our survey. In addition, Hall et al. (1996) looked at the methodological quality of surgical clinical trials.[21] They reported that less than 50% of the studies commented on potential bias in the assessment of the outcome, adequately described the randomization technique, or commented on sample size calculation.[21] Peng et al. (2006) conducted a review on published epidemiological papers to assess the reproducibility of epidemiological research.[22] They found that 30% of the publications did not report the implementation of the statistical analysis. Begley et al. (2015) commented that there is a general problem of reproducibility of study results, in particular in preclinical studies.[23] This goes in line with the problems of study protocols and designs reported by Ioannidis et al. (2014).[17]

A total of 57 (66%) of the contacted 86 persons participated and fulfilled the inclusion criteria. This corresponds to a high participation proportion. However, it remains unknown whether all potential participants were truly identified and contacted.

A further limitation of our survey is that it does not provide an overview of the objectively measured quality and completeness of biostatistical aspects of study protocols but that it only refers to the subjective, individual impression of the statistical members of the ethics committees. An objective rating of the study protocols, however, was not possible due to data and privacy protection issues, as this would have required screening of the submitted study documents. Moreover, objective quality measures are difficult to define, because as the meaning of "adequate" quality might differ considerably. The subjective ratings of completeness and correctness are subject to inter- and intra-rater variability. Therefore, the results should rather be interpreted as a rough indication and not as definite numbers.

As a third limitation, we consider the fact that the survey could not assess recent issues added in an addendum to the ICH E9 guideline.[24] It presents a structured framework to link trial objectives to a

suitable trial design and tools for estimation and hypothesis testing. This framework introduces the concept of an estimand, translating the trial objective into a precise definition of the treatment effect that is to be estimated. It also aims to facilitate the dialogue between disciplines involved in a clinical trial.

Even in view of these limitations, the survey clearly indicates the need for basic and advanced statistical trainings and guidance for medical researchers. All medical faculties in Germany have established biostatistical units providing consulting services. However, this does not seem to be sufficient to enable medical researchers to develop protocols, which cover statistical issues adequately. Reasons could be that medical researchers undervalue the impact of an appropriate biometrical planning on the quality and validity of medical studies. In personal discussions with participants of this survey, several of them reported frequent examples where the statistical analysis strategy in study protocols is addressed with a single general sentence like "The data are analyzed with valid statistical methods." This does not only indicate a lack of statistical knowledge but also a lack of awareness that statistical methods have an important impact on the validity of medical research.

The survey also gauges the range of methodological challenges encountered by biostatisticians being member of a medical ethics committee. Often a biostatistician is the last methodological sentinel before a study is implemented in a clinical setting. In order to involve more biostatisticians in medical ethics committees, there is a need to provide support to enable them to adequately discharge their responsibilities. Unfortunately, the survey did not check how the remaining members react on revision requests by biostatisticians and if these requests are adequately addressed before the final vote of the medical ethics committee on the criticized study protocol. Moreover, we did not formally assess if biostatisticians, who are members of medical ethics committees, formulate their requests in comparable detail and persistence. Due to own experiences and based on narrative reports, we suspect that biostatistical concerns cannot be easily communicated to and understood by the non-statistical members of the medical ethics committee. Thus, there is a need to establish a better communication, which allows expressing biostatistical concerns in a convincing easily understandable language.

It is therefore time to communicate the general importance of statistics for medical research. This includes the establishment of guidelines for protocol writing and templates like the SPIRIT Statement, which also handles statistical input to a protocol.[25,26] Furthermore, the implementation of reporting guidelines like STROBE should be made more popular.[27] Moreover, the development of specific trainings and guidance on how to address specific statistical challenges is required. Finally, national standards for the tasks of a biostatistician as a member of a medical ethics committee must be formulated.[28,29]

## References

1. General Assembly of the World Medical Association (2014). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *The Journal of the American College of Dentists* 81: 14.
2. Doppelfeld E, Hasford J (2019). [Medical research ethics committees in the Federal Republic of Germany: establishment and integration into medical research]. [in German] *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 62: 682-689.

3. Buchner B, Hase F, Borchers D, Pigeot I (2019). [Tasks, regulations, and functioning of ethics committees]. [in German] *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 62: 690-696.
4. Gelfond JA, Heitman E, Pollock BH, Klugman CM (2011). Principles for the ethical analysis of clinical and translational research. *Stat Med* 30:2785-2792.
5. Schlattmann P, Scherag A, Rauch G, Mansmann U (2019). [The role of biostatistics in institutional review boards]. [in German] *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 62: 751-757.
6. International Conference on Harmonization (1999). ICH E9 Guideline. Statistical principles for clinical trials. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf (accessed 07. October 2019)
7. Raspe H, Hüppe A, Strech D, JT (2012). Empfehlungen zur Begutachtung klinischer Studien durch Ethik-Kommissionen. Deutscher Ärzte-Verlag.
8. Victor N (1999). Prüfung der wissenschaftlichen Qualität und biometriespezifischer Anforderungen durch die Ethikkommission. *Medizinrecht* 17: 408-412.
9. Bundesgesetzblatt (2016). Viertes Gesetz zur Änderungen arzneimittelrechtlicher und anderer Vorschriften, 20. Dezember 2016, Teil I Nr.63. https://www.bgbl.de/xaver/bgbl/start.xav?start=%2F%2F*[%40attr_id%3D%27bgbl116s3048.pdf%27]#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl116s3048.pdf%27%5D__1570442107702 (accessed 07. October 2019)
10. Gesetz über den Verkehr mit Arzneimitteln (Arzneimittelgesetz−AMG). https://www.gesetze-im-internet.de/amg_1976/ (accessed 07. October 2019)
11. Gesetz über Medizinprodukte (Medizinproduktegesetz−MPG). http://www.gesetze-im-internet.de/mpg/index.html (accessed 07. October 2019)
12. Lanzerath, D. (2019). [European ethics committees in transition: challenges of new requirements]. [in German] *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz* 62: 697-705.
13. Pigeot I, Foraita R, Frömke C, Haller B, Hoffmann VS, Zapf A (2019). [Ethical consideration of studies involving human subjects outside the regulatory framework: not mandatory, but of high relevance]. [in German] *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 62: 722-728. Erratum in: *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2019; 62: 1276-1277.
14. Rauch G, Röhmel J, Gerß J, Scherag A, Hofner B (2019). [Current challenges in the assessment of ethical proposals-aspects of digitalization and personalization in the healthcare system]. [in German] *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 62: 758-764.
15. Good P, Hardin J (2009). Common errors in statistics (and how to avoid them). Wiley.
16. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwarz Lm, Woloshin S (2008). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest* 8: 53-96.
17. Ioannidis JP, Greenland S, Hlatky MA, et al. (2014). Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet* 383: 166-175.
18. Wang MQ, Yan AF, Katz RV (2018). Researcher requests for an appropriate analysis and reporting: A U.S. survey of consulting biostatisticians. *Ann Intern Med* 169: 554-558.
19. Clark T, Berger U, Mansmann U (2013): Sample size determinations in original research protocols for randomised clinical trials submitted to UK research ethics committees: review. *British Medical Journal* 346:f1135.
20. Kilkenny C, Parsons N, Kadyszewski E, Festing MF et al. (2009). Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PloS One* 4: e7824.

21. Hall JC, Mills B, Nguyen N, Hall JL (1996). Methodologic standards in surgical trials. *Surgery* 119: 466-472.
22. Peng RD, Dominici F, Zeger SL (2006). Reproducible epidemiologic research. *American Journal of Epidemiology* 163: 783-789.
23. Begley CG, Buchan AM, Dirnagl U (2015). Robust research: Institutions must do their part for reproducibility. *Nature News* 525: 25.
24. International Conference on Harmonization (2017) ICH E9 (R1) Addendum. ICH E9 (R1) Addendum on estimands and sensitivity5 analysis in clinical trials to the guideline on statistical 6 principles for clinical trials https://www.ema.europa.eu/en/documents/scientific-guideline/draft-ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical_en.pdf (accessed 09. September 2019)
25. Chan AW, Tetzlaff JM, Altman DG, Laupacis A et al. (2013). SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Annals of Internal Medicine* 158: 200-207.
26. Clark T, Davies H, Mansmann U (2014): Five questions that need answering when considering the design of clinical trials. *Trials* 15: 286.
27. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, & Vandenbroucke JP (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Annals of Internal Medicine* 147: 573-577.
28. Zapf A, Huebner M, Rauch G, Kieser M (2018). What makes a biostatistician? *Statstics in Medicine* 38:695-701.
29. Gerlinger C, Edler L, Friede T et al. (2012). Considerations on what constitutes a 'qualified statistician' in regulatory guidelines. *Statistics in Medicine* 31:1303-1305.


**Author's contribution**

- Geraldine Rauch developed the research idea, designed the survey and wrote the manuscript.
- Lorena Hafermann implemented the online survey, performed the analyses and reviewed the manuscript.
- Ulrich Mansmann helped to develop the research idea, reviewed the survey, reviewed the manuscript.
- Iris Pigeot helped to develop the research idea, reviewed the survey, reviewed the manuscript.

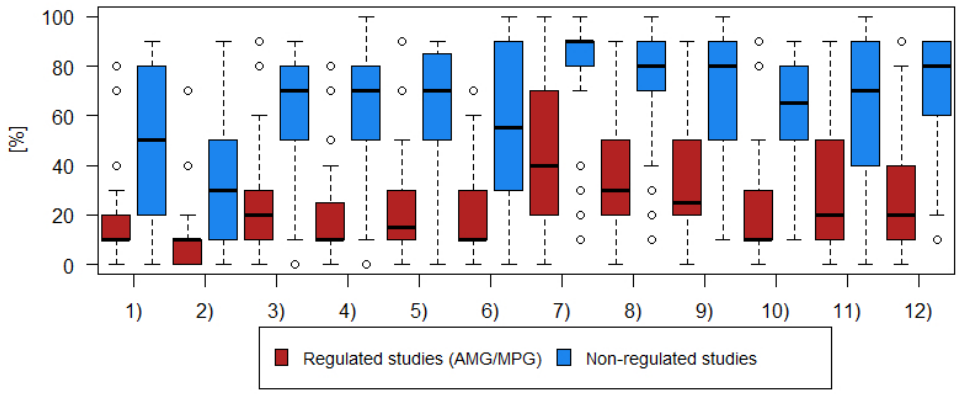
**A data sharing statement**

The original dataset of the survey is available from the Dryad repository, https://datadryad.org/stash/share/VjofDJkoUtqIjjaJQ84O7FB1W5cJu78g04cNey044no.

**Figure Legends**

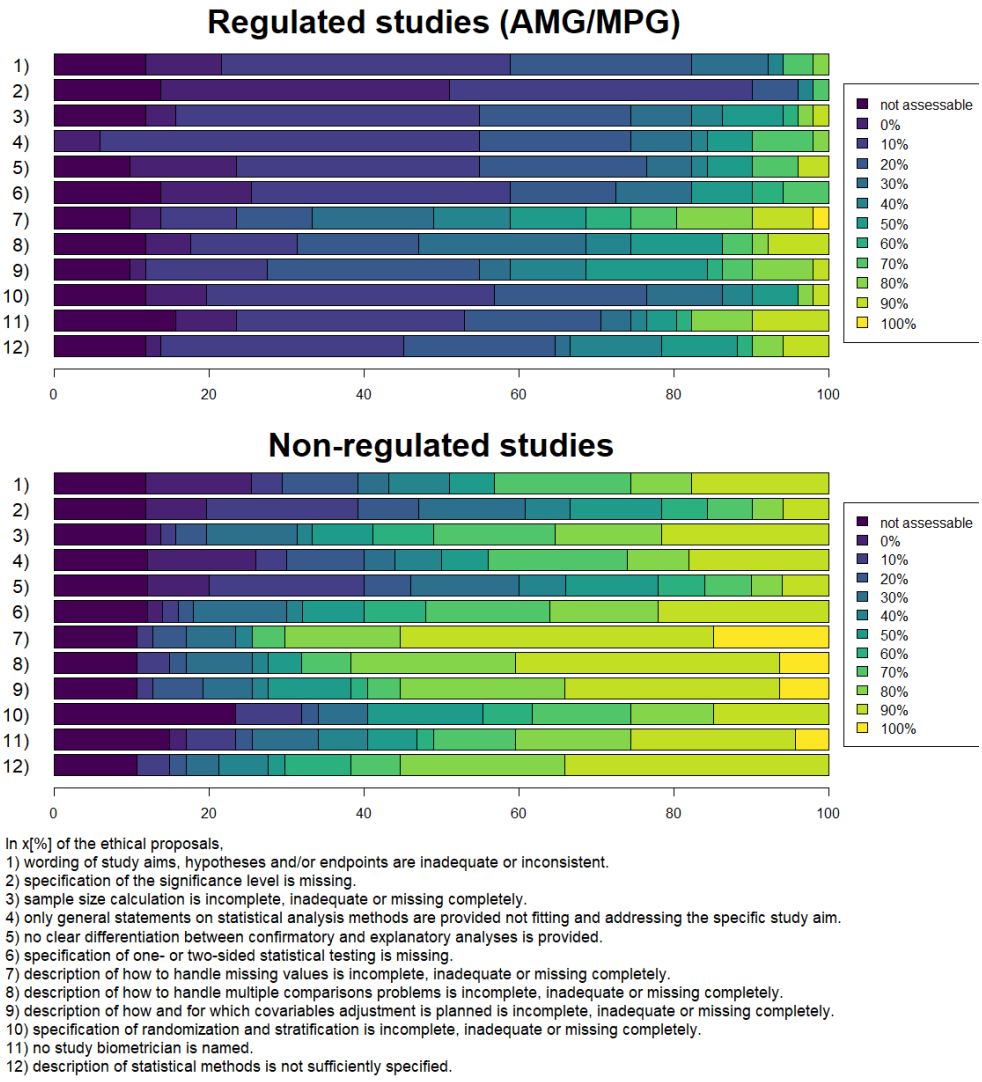
Figure 1. Grouped boxplots for completeness and correctness of 12 biostatistical aspects for regulated and non-regulated studies.

Figure 2. Staggered bar plots for completeness and correctness of 12 biostatistical aspects differing between regulated and non-regulated studies.




In x[%] of the ethical proposals,
1) the wording of the study aims, the hypotheses and/or the endpoints is inadequate or inconsistent.
2) the specification of the significance level is missing.
3) the sample size calculation is incomplete, inadequate or missing completely.
4) only general statements on statistical analysis methods are provided not fitting and addressing the specific study aim.
5) no clear differentiation between confirmatory and explanatory analyses is provided.
6) the specification of one- or two-sided statistical testing is missing.
7) the description of the handling with missing values is incomplete, inadequate or missing completely.
8) the description of the handling with multiple testing problems is incomplete, inadequate or missing completely.
9) the description of how and for which covariables adjustment is planed is incomplete, inadequate or missing completey.
10) the specification of randomisation and stratification, is incomplete, inadequate or missing completely.
11) no study biometrician is named.
12) the description of the statistical methods is not sufficiently specified.

Grouped boxplots for completeness and correctness of 12 biostatistical aspects for regulated and non-regulated studies.




Staggered bar plots for completeness and correctness of 12 biostatistical aspects differing between regulated and non-regulated studies.

# Original questions of the survey in in English translation

**1) Do you currently work or have you been working as biometrician/statistician in a medical ethics commission?**

- yes
- no

**2) In which kinds of medical ethic committees have you been working?**

- medical ethic committee of a medical faculty of a university
- medical ethic committee of a State Chamber of Physicians (Landesärztekammer)
- others

**3) Which German federal state is the medical ethics committees located to which you were affiliated?**

- Baden-Württemberg
- Bavaria
- Berlin
- Brandenburg
- Bremen
- Hamburg
- Hesse
- Mecklenburg Western Pomerania
- Lower Saxony
- Northrhine - Westphalia
- Rhineland Palatinate
- Saarland
- Saxony - Anhalt
- Schleswig Holstein
- Thuringia
- Saxony

**4) How many study proposals (amendments and annual reports excluded) do you review on average per year?**

- up to 50
- 51 - 100
- 101 - 150
- 151 - 200
- more than 200

**5) Which kinds of studies do you review as member of a medical ethics committee?**

- Studies regulated by AMG (German Drug Act) or MPG (German Medical Products Act)
- Studies in a non-regulated setting
- Both

**6) Do you in general have the impression that the biostatistical quality of ethical proposals for studies regulated by AMG/MPG differs compared to proposals in a non-regulated setting?**

- Yes, ethical proposals for studies regulated by AMG/MPG have a higher biostatistical quality on average
- Yes, ethical proposals for studies in a non-regulated setting have a higher biostatistical quality on average
- No, the biostatistical quality does not differ on average
- Not specified

**7) Please complete the following statements related to biostatistical problems in ethical proposals for studies regulated by AMG/MPG through specification of approximate percentages according to your subjective assessment.**

**Part 1**

- In x[%] of the ethical proposals, wording of study aims, hypotheses and/or endpoints are inadequate or inconsistent.
- In x[%] of the ethical proposals, specification of the significance level is missing.
- In x[%] of the ethical proposals, sample size calculation is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals, only general statements on statistical analysis methods are provided not fitting and addressing the specific study aim.
- In x[%] of the ethical proposals, no clear differentiation between confirmatory and explanatory analyses is provided.
- In x[%] of the ethical proposals, specification of one- or two-sided statistical testing is missing.

**8) Part 2**

- In x[%] of the ethical proposals, description of how to handle missing values is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals, description of how to handle multiple comparisons problems is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals, description of how and for which covariables adjustment is planned is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals, specification of randomization and stratification is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals no study biometrician is named.
- In x[%] of the ethical proposals, description of statistical methods is not sufficiently specified.

**9) Please complete the following statements related to biostatistical problems in ethical proposals for non-regulated studies through specification of approximate percentages according to your subjective assessment.**

**Part 1**

- In x[%] of the ethical proposals, wording of study aims, hypotheses and/or endpoints are inadequate or inconsistent.
- In x[%] of the ethical proposals, specification of the significance level is missing.
- In x[%] of the ethical proposals, sample size calculation is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals, only general statements on statistical analysis methods are provided not fitting and addressing the specific study aim.

- In x[%] of the ethical proposals, no clear differentiation between confirmatory and explanatory analyses is provided.
- In x[%] of the ethical proposals, specification of one- or two-sided statistical testing is missing.

**10) Part 2**

- In x[%] of the ethical proposals, description of how to handle multiple comparisons problems is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals, description of how and for which covariables adjustment is planned is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals, specification of randomization and stratification is incomplete, inadequate or missing completely.
- In x[%] of the ethical proposals no study biometrician is named.
- In x[%] of the ethical proposals, description of statistical methods is not sufficiently specified.

**11) How do you consider the need for additional training in the following biostatistical areas for investigators submitting protocols to medical ethics committees? (low, middle, high, not assessable)**

- Study design
- Wording of the study aims, hypotheses and/or endpoints
- Sample size calculation
- Differentiation between confirmatory and exploratory analyses
- Handling of missing values
- Description of the statistical analysis
- Multiple comparisons problems
- Adjustment for covariables
- Randomization/stratification

**12) Did you enjoyed our survey?**

- yes
- no

## Checklist "Standards for Reporting Qualitative Research (SRQR)"*

This manuscript describes a quantitative survey. None of the provided checklists perfectly fits this type of empirical study. We felt that the checklist "Standards for Reporting Qualitative Research (SRQR)*" in many aspects addresses our type of study, although this is not a qualitative survey.

### Standards for Reporting Qualitative Research (SRQR)*

http://www.equator-network.org/reporting-guidelines/srqr/

| | Page/line no(s). |
|---|---|
| **Title and abstract** | |
| **Title** - Concise description of the nature and topic of the study Identifying the study as qualitative or indicating the approach (e.g., ethnography, grounded theory) or data collection methods (e.g., interview, focus group) is recommended | Page 1 |
| **Abstract** - Summary of key elements of the study using the abstract format of the intended publication; typically includes background, purpose, methods, results, and conclusions | Page 2 |
| **Introduction** | |
| **Problem formulation** - Description and significance of the problem/phenomenon studied; review of relevant theory and empirical work; problem statement | Page 4 |
| **Purpose or research question** - Purpose of the study and specific objectives or questions | Page 4 |
| **Methods** | |
| **Qualitative approach and research paradigm** - Qualitative approach (e.g., ethnography, grounded theory, case study, phenomenology, narrative research) and guiding theory if appropriate; identifying the research paradigm (e.g., postpositivist, constructivist/ interpretivist) is also recommended; rationale** | Page 5 |
| **Researcher characteristics and reflexivity** - Researchers' characteristics that may influence the research, including personal attributes, qualifications/experience, relationship with participants, assumptions, and/or presuppositions; potential or actual interaction between researchers' characteristics and the research questions, approach, methods, results, and/or transferability | Page 5 |
| **Context** - Setting/site and salient contextual factors; rationale** | Page 5 |
| **Sampling strategy** - How and why research participants, documents, or events were selected; criteria for deciding when no further sampling was necessary (e.g., sampling saturation); rationale** | Page 6 |
| **Ethical issues pertaining to human subjects** - Documentation of approval by an appropriate ethics review board and participant consent, or explanation for lack thereof; other confidentiality and data security issues | Page 7 |

| | |
|---|---|
| **Data collection methods** - Types of data collected; details of data collection procedures including (as appropriate) start and stop dates of data collection and analysis, iterative process, triangulation of sources/methods, and modification of procedures in response to evolving study findings; rationale** | Page 5 |
| **Data collection instruments and technologies** - Description of instruments (e.g., interview guides, questionnaires) and devices (e.g., audio recorders) used for data collection; if/how the instrument(s) changed over the course of the study | Page 5-6 |
| **Units of study** - Number and relevant characteristics of participants, documents, or events included in the study; level of participation (could be reported in results) | Page 6 |
| **Data processing** - Methods for processing data prior to and during analysis, including transcription, data entry, data management and security, verification of data integrity, data coding, and anonymization/de-identification of excerpts | Page 6 |
| **Data analysis** - Process by which inferences, themes, etc., were identified and developed, including the researchers involved in data analysis; usually references a specific paradigm or approach; rationale** | Page 7 |
| **Techniques to enhance trustworthiness** - Techniques to enhance trustworthiness and credibility of data analysis (e.g., member checking, audit trail, triangulation); rationale** | Page 6 |

**Results/findings**

| | |
|---|---|
| **Synthesis and interpretation** - Main findings (e.g., interpretations, inferences, and themes); might include development of a theory or model, or integration with prior research or theory | Pages 7-11 |
| **Links to empirical data** - Evidence (e.g., quotes, field notes, text excerpts, photographs) to substantiate analytic findings | x |

**Discussion**

| | |
|---|---|
| **Integration with prior work, implications, transferability, and contribution(s) to the field -** Short summary of main findings; explanation of how findings and conclusions connect to, support, elaborate on, or challenge conclusions of earlier scholarship; discussion of scope of application/generalizability; identification of unique contribution(s) to scholarship in a discipline or field | Pages 11-13 |
| **Limitations** - Trustworthiness and limitations of findings | Pages 12-13 |

**Other**

| | |
|---|---|
| **Conflicts of interest** - Potential sources of influence or perceived influence on study conduct and conclusions; how these were managed | Page 3 |
| **Funding** - Sources of funding and other support; role of funders in data collection, interpretation, and reporting | Page 3 |

*The authors created the SRQR by searching the literature to identify guidelines, reporting standards, and critical appraisal criteria for qualitative research; reviewing the reference lists of retrieved sources; and contacting experts to gain feedback. The SRQR aims to improve the transparency of all aspects of qualitative research by providing clear standards for reporting qualitative research.

**The rationale should briefly discuss the justification for choosing that theory, approach, method, or technique rather than other options available, the assumptions and limitations implicit in those choices, and how those choices influence study conclusions and transferability. As appropriate, the rationale for several items might be discussed together.