

QUALITY OF SAMPLE SIZE ESTIMATION IN TRIALS OF MEDICAL DEVICES: HIGH-RISK DEVICES FOR NEUROLOGICAL CONDITIONS AS EXAMPLE

Britta Olberg

Medical Consultancy Department, German Federal Joint Committee
Department of Health Care Management, Berlin University of Technology
b.olberg@campus.tu-berlin.de

Matthias Perleth

Katja Felgentraeger

Sandra Schulz

Medical Consultancy Department, German Federal Joint Committee

Reinhard Busse

Department of Health Care Management, Berlin University of Technology

Background: The aim of this study was to assess the quality of reporting sample size calculation and underlying design assumptions in pivotal trials of high-risk medical devices (MDs) for neurological conditions.

Methods: Systematic review of research protocols for publicly registered randomized controlled trials (RCTs). In the absence of a published protocol, principal investigators were contacted for additional data. To be included, trials had to investigate a high-risk MD, registered between 2005 and 2015, with indications stroke, headache disorders, and epilepsy as case samples within central nervous system diseases. Extraction of key methodological parameters for sample size calculation was performed independently and peer-reviewed.

Results: In a final sample of seventy-one eligible trials, we collected data from thirty-one trials. Eighteen protocols were obtained from the public domain or principal investigators. Data availability decreased during the extraction process, with almost all data available for stroke-related trials. Of the thirty-one trials with sample size information available, twenty-six reported a predefined calculation and underlying assumptions. Justification was given in twenty and evidence for parameter estimation in sixteen trials. Estimates were most often based on previous research, including RCTs and observational data. Observational data were predominantly represented by retrospective designs. Other references for parameter estimation indicated a lower level of evidence.

Conclusions: Our systematic review of trials on high-risk MDs confirms previous research, which has documented deficiencies regarding data availability and a lack of reporting on sample size calculation. More effort is needed to ensure both relevant sources, that is, original research protocols, to be publicly available and reporting requirements to be standardized.

Keywords: Medical devices, Randomized controlled trials, Sample size calculation, Neurological

The market for high-risk medical devices (MDs) and implants has shown significant growth over the past decade (1). One of the driving market segments behind this development is that of neurostimulation devices. This is due to high demand for invasive and noninvasive treatment options for neurological conditions. Demand for such options has been growing, in part, by increasing incidence and government funding toward research into such diseases (2). For example, Storz et al. found that central nervous disease (CNS) disorders are among a group of disease areas with highly active research (3).

When it comes to decisions regarding the reimbursement of new procedures involving MDs within a health insurance system, key players such as regulatory bodies, funding agencies, and patients demand results from clinical trials, such as randomized controlled trials (RCTs), with high scientific validity and reliability. It is essential that these stakeholders have confidence in the research and are aware of the strengths and

weaknesses of the methodologies used for trial implementation (4). For example, because the effect size strongly influences a decision, the involved parties should be aware of the relevance of the sample size that is needed to demonstrate such an effect according to a prespecified power calculation. The sample size, in turn, is a decisive part of RCT planning (5). Hence, it is crucial that the reporting of RCTs is transparent with respect to sample size and, in particular, underlying assumptions regarding the anticipated treatment effect.

Despite its importance, previous studies repeatedly found a lack of data availability and reporting of the sample size methodology including justification for the values of a priori estimates. All of them conclude that sample sizes are poorly reported, are erroneous, and often based on inaccurate assumptions regarding the expected event rate (4;6–9). Nevertheless, a considerable number of approaches and guidelines for reporting in RCTs exists in the literature (10;11). One important example is the revised consolidated standards for reporting trials

(CONSORT) statement, which has been developed to improve suboptimal reporting of RCTs (4;12).

In addition, leading medical journals (e.g. BMJ, Lancet) increasingly only publish trials in accordance with the CONSORT recommendations and require submitted study publications to be accompanied by original research protocols (13;14). Although such guidelines and preconditions clearly specify and encourage standardized reporting of the sample size methodology, deficiencies have still been widely documented in the literature; even years after these guidelines have first been introduced. Therefore, greater transparency and better justification of the sample size estimation is recommended to promote the early detection of shortcomings in study design (4;6;7;15–18).

However, analyses typically focus on pharmaceutical studies and have been limited for example to specific journals (15, 18), a certain research ethics committee (7;16), a specific study design (17), or only one study registry (8). To our knowledge, no such systematic analyses have been performed for trials of high-risk MDs. Therefore, we formulate two hypotheses regarding the documentation of sample size calculation in trials testing MDs. First, the rationale for underlying assumptions is incomplete or non-existent in most cases. Second, when sample size estimation is reported, parameters are derived from scientific evidence of poor quality.

To test these hypotheses, we conducted a systematic review aiming at assessing the completeness of reporting sample size calculation and the underlying quality of evidence for sample size calculation. Complete study protocols are reliable sources of information with a key role in reducing bias by documenting a prespecified blueprint with respect to the conduct and analyses of trials (7). We, therefore, primarily focused on the identification of original research protocols.

METHODS

Trial Search

Our initial search was divided into three steps:

Preliminary search. In the run-up to our systematic examination of registered trials, we performed an intensive exploratory search using ClinicalTrials.gov to identify the most frequently investigated indications relying on MDs for neurological conditions. Based on these results, stroke, epilepsy, and headache disorders, were selected as case samples for further analyses (see Supplementary Tables 1 and 2). As these three indications represent a broad spectrum of neurological conditions (e.g., chronic versus acute, range of pharmaceutical and nonpharmaceutical interventions available, high incidence and prevalence), they enhance the applicability of our results.

Main search. The second step was a focused search for these three indications using the two trial resources ClinicalTrials.gov and the International Clinical Trials Registry Platform (ICTRP) of

the World Health Organization (WHO). We chose these two clinical trial databases to ensure that the included trials are as representative as possible. ClinicalTrials.gov represents the largest and most widely used registry of clinical trials. However, its focus is on studies conducted in the United States of America (USA). To mitigate any sample bias, we included the WHO/ICTRP database, which aggregates trials from many regional registries around the world, in particular Europe. To systematically identify all relevant registered clinical trials for MDs aimed at the selected indications, search criteria such as the indication, the type of study and the trial phase were applied. Only trials registered between 2005 and 2015 were included, as legal and regulatory frameworks have not changed significantly in this period. The detailed search strategy of both databases is provided as supplementary material of this article (see Supplementary Table 3).

Identification of full study protocols. The third step was to identify study protocols including extra material such as statistical appendices for the previously identified trials. In case no study protocol was available, we searched for the corresponding study article. Additionally, and as a last resort, we attempted to contact the principal investigators or trial sponsors to ask for provision of the missing protocol or additional data such as the statistical analysis plan.

Trial Selection

Inclusion and exclusion criteria were determined a priori to guide trial screening and selection to ensure identification of the most relevant body of literature with respect to the research question. Specifically, we included all interventional, randomized, comparative phase II, III, and IV trials of MDs defined according to the EU Directives 90/385/EEC on active implantable MDs and 93/42/EEC on MDs with a risk class of at least IIB (19).

We excluded trials with study designs other than RCTs, trials that investigated other healthcare interventions (e.g., pharmaceuticals), trials that focused on MDs of risk classes below IIB, and/or trials that did not match the preset indications. A tabular presentation of all inclusion and exclusion criteria used for data selection is given in the Supplementary Table 4.

Database Generation and Data Extraction

General characteristics of the trials from the databases were extracted into a spreadsheet, including, for example, disease area, study design, type, and the risk class of the MD. For each trial, the protocol (including statistical appendices, amendments, etc.) and in the absence of a protocol the corresponding publication or any additional information from the principal investigators or sponsors was reviewed. We then extracted data with respect to key methodological items relevant for assessing the sample size calculation. These methodological extraction items were defined based on the literature and discussions

among the research team. The extraction followed a hierarchical order. We started with the availability of sample size calculations, followed by design assumptions, their justification, and finally the underlying scientific evidence. Any of the subsequent items were only relevant, if the preceding items were answered positively. We did not assess the appropriateness of the methods used for sample size estimation. The definitions of the extraction items used are as follows:

1. Reporting Sample Size Calculation. This item is a binary “yes/no” option and refers to the reporting in the document only.
2. Reporting Design Assumptions. This item is a binary “yes/no” option and refers to the reporting in the document only. The term “design assumptions” relates to the main elements used for calculation of the sample size guided by the CONSORT recommendations (20). These include: (i) the estimated outcomes in each group, which suggests the clinically important target difference between the intervention groups; (ii) the α (type I) error level; (iii) the statistical power (or the β (type II) error level); and (iv) for continuous outcomes, the standard deviation of the measurements. To be assessed as “reported” it was presumed that these elements were presented.
3. Reporting Justification. This item is a binary “yes/no” option and refers to the reporting in the document only. Justification required an explanation for the selection of the value(s) for the sample size calculation including a discussion why these were assessed as reasonable for the study.
4. Reporting Evidence on which Assumptions Are Based. For this item, multiple sources were possible including:
 - results from a previous/preliminary trial (RCT, observational data, etc.),
 - results from a review (systematic or narrative),
 - other (interim analyses, databases, expert opinions, etc.).

All extractions were carried out by one researcher and independently checked by another. Discrepancies were discussed and a final data pool was consented. The detailed data sheets for all three indications will be provided on request.

RESULTS

Study Pool

Our search yielded a total of 1,074 publicly registered trials, with 616 remaining after removal of duplicates. Subsequent screening for relevance, using our predefined in- and exclusion criteria (see Supplementary Table 4), resulted in the exclusion of further 543 trials. Seventy-three trials (forty-nine on stroke and twelve studies each on headache disorders and epilepsy, respectively) fulfilled the inclusion criteria and were included in our sample (see Figure 1).

Data Pool for Qualitative Synthesis

Documents Obtained from Registry Databases. Overall, for twelve of the seventy-three trials, full study protocols were available in the registry databases. These protocols were exclusively from the stroke sub-sample. For sixty-one trials (thirty-seven for stroke and twelve for each headache disorders and epilepsy, respec-

tively), no full study protocols could be obtained (see Figure 2). For nine of the sixty-one trials, a corresponding publication describing the study methods was identified in the registry databases (five for stroke, two for epilepsy, and two for headache disorders) (see Figure 2).

Information Gathered from Experts. Regarding the sixty-one trials with no publicly available study protocol in the registry databases, we contacted a total of seventy-six principal investigators and/or sponsors to obtain additional information on fifty-seven studies. For four studies, no current contact information was available.

Attempts to contact these experts were unsuccessful for more than half of the studies (34 studies). However, of the twenty-three responses received, all regarding different studies, fourteen were judged as useful and nine were negative (i.e., refusal to share information, mainly because of confidentiality or time constraints). Of the useful responses, six yielded additional full study protocols (five for stroke and one for epilepsy) (see Figure 2). The other eight contained additional information from principal investigators and/or sponsors themselves, such as protocol extracts or descriptions regarding the sample size estimation. Based on this additional information, two trials were subsequently excluded, due to the respondents' disclosure about the exact study design (e.g., a feasibility study with no planned sample size calculation did not match our inclusion criteria) (see Figure 2). No additional supplementary publications were provided.

Final Data Pool. Overall, a total of seventy-one trials fulfilled our inclusion criteria. From these trials, we deduced our final data pool for qualitative analyses, consisting of thirty-one trials for which sufficient data regarding the sample size estimation were available. Specifically, in eighteen cases corresponding full research protocols could be obtained. For another eight, we relied on publications. Finally, for five studies, experts provided sufficient additional information to justify consideration in subsequent analyses. Note that, despite an initial number of nine publications, one publication was disregarded in favor of a study protocol provided by an expert upon request. Moreover, for one trial we did obtain additional information from an expert, but could not consider it in subsequent analyses due to lack of detail (see Figure 2).

Overall Data Availability

Data availability decreased along the hierarchical data extraction process (see Figure 3). As documented above, less than half of the included trials (31 of 71), provided any data for further analysis (i.e., study protocol, publication, or additional data provided by means of experts). Eighty percent of the available data stemmed from stroke trials (twenty-five trials). Data accessibility decreased further along the process, as only half of these trials (sixteen of thirty-one) reported any scientific evidence on

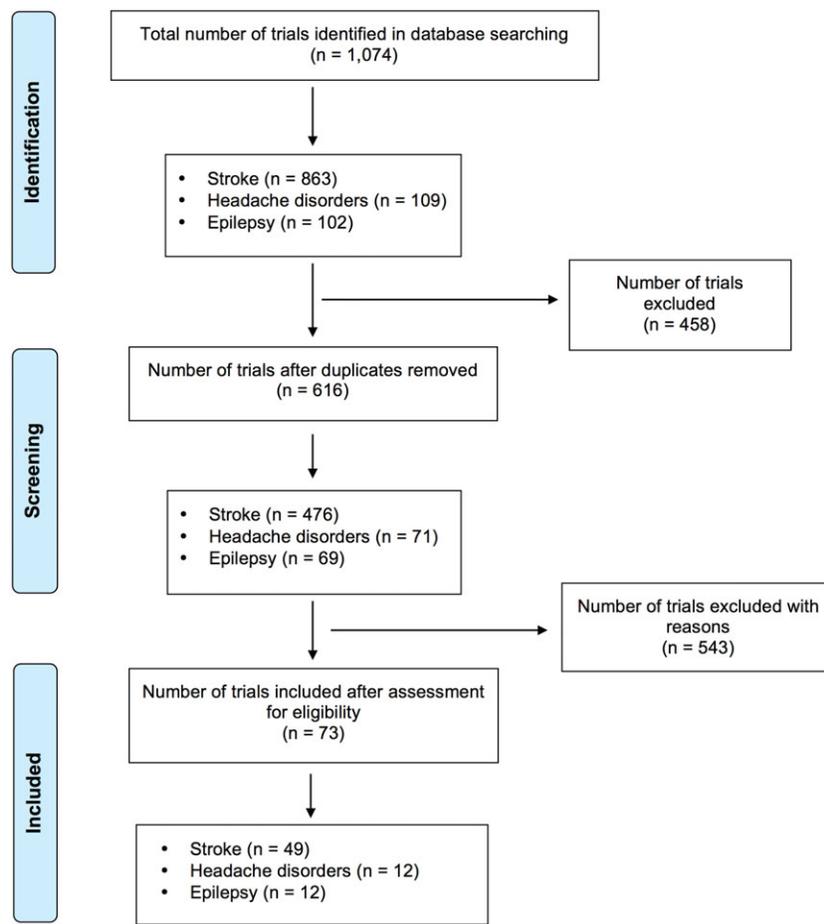


Figure 1. Trial selection.

the assumptions they were based on. At this point, data only came from stroke related trials. No data were available for the indications epilepsy and headache disorders.

Reporting of Sample Size Methodology

Among the thirty-one trials for which data were available, twenty-six reported an a priori sample size calculation. Five trials did not provide any information. All of the twenty-six trials containing information about sample size estimation also stated underlying assumptions. Of the trials that reported assumptions, twenty justified the respective assumptions, whereas six trials did not. Within the twenty trials that provided a justification, sixteen reported evidence from which values a priori estimates were derived. The remaining four trials did not provide any reasoning supported by data.

Evidence behind the justifications was only available for stroke trials. A single trial could refer to multiple study types and studies. For example, a trial could refer to four observational studies, of which one was retrospective and three were prospective. Our results were as follows: where justified, parameters were most often set based on previous/preliminary research (in fourteen of sixteen trials), coming from RCTs (in

six trials), single-arm studies (in two trials) and observational data (in six trials). The observational data used by the six trials stemmed from five retrospective and three prospective studies. In three trials, the study design used for estimation was not assessable due to a lack of detailed information. Moreover, in three of sixteen trials, approximations were also based on reviews. Two reviews were nonsystematic or narrative articles; one review was conducted systematically. Nevertheless, the studies included in the systematic review were predominately retrospective. Finally, in three trials, variables were deduced from results of another type of evidence such as an interim or a database analysis. Details of sample size reporting are given in Figure 4.

DISCUSSION

Our results are in line with previous research documenting poor reporting and, for the first time, highlight deficiencies within trials testing high-risk MDs.

In this systematic review of seventy-one trials from the past 10 years in the area of CNS diseases, findings are twofold. On the one hand, the review primarily shows the generally poor data availability. On the other hand, it illustrates the opacity of

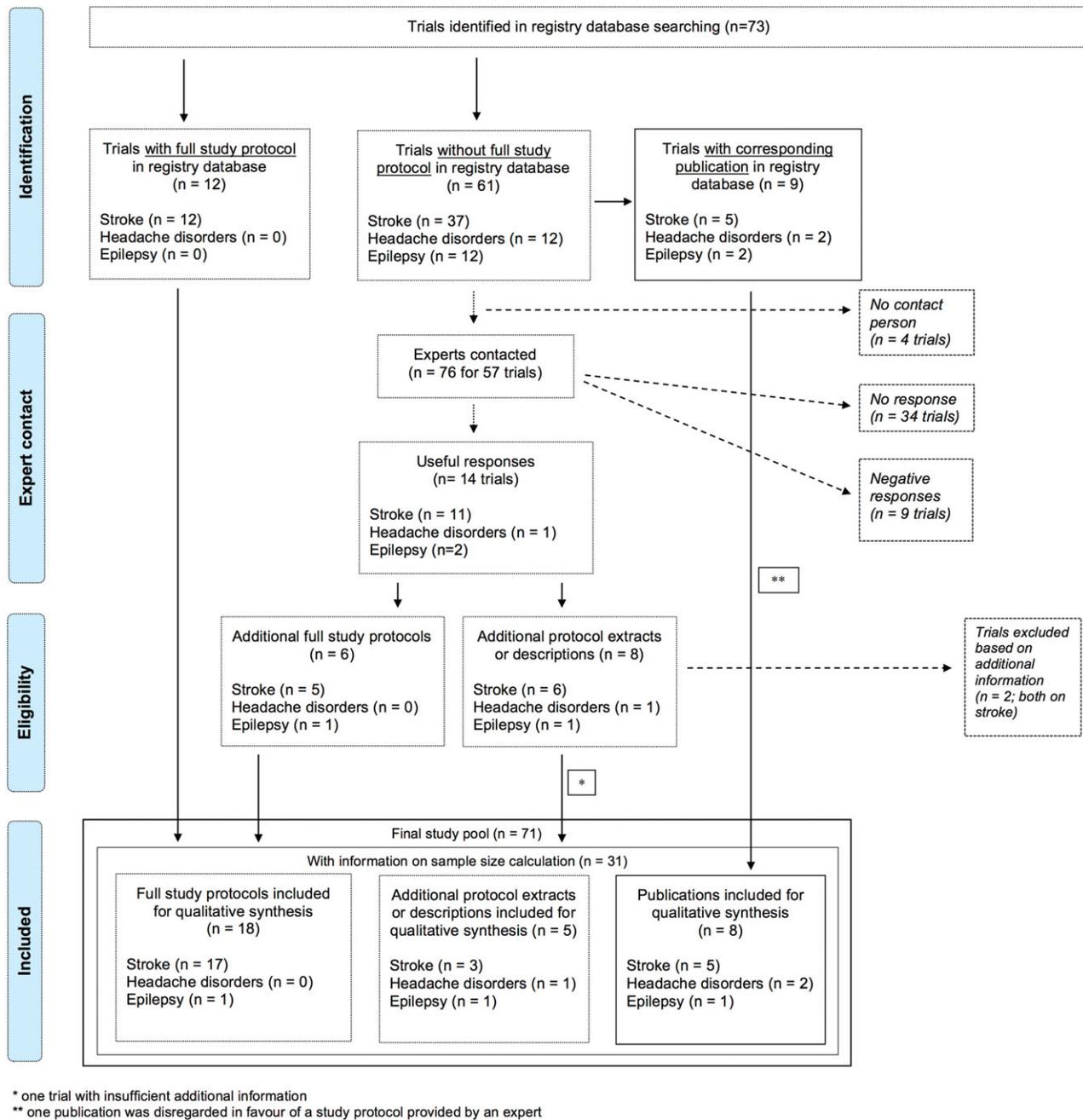


Figure 2. Data pool for qualitative analysis.

sample size methodology, especially when it comes to scientific evidence supporting parameter estimation. More precisely, for fewer than half of the included trials we were able to find data for our analysis. Almost all information remaining after our extraction process stemmed from stroke trials, which shows the need for more research on other indications in which high-risk MDs are involved.

Although additional data obtained from experts or a corresponding paper added some information that we used to fill any gaps, we recognized that protocols are the most valuable source

for information on study designs. However, of the trials for which data were available, only half reported evidence underlying parameter estimation. The remaining trials did not report any evidence or lacked detail, making any assessment impossible. When parameter estimates were justified based on clinical evidence these came most often from previous trials including RCTs. However, at least a comparable number of the trials extrapolated their sample size parameter from observational data, which represented predominantly retrospective study designs.

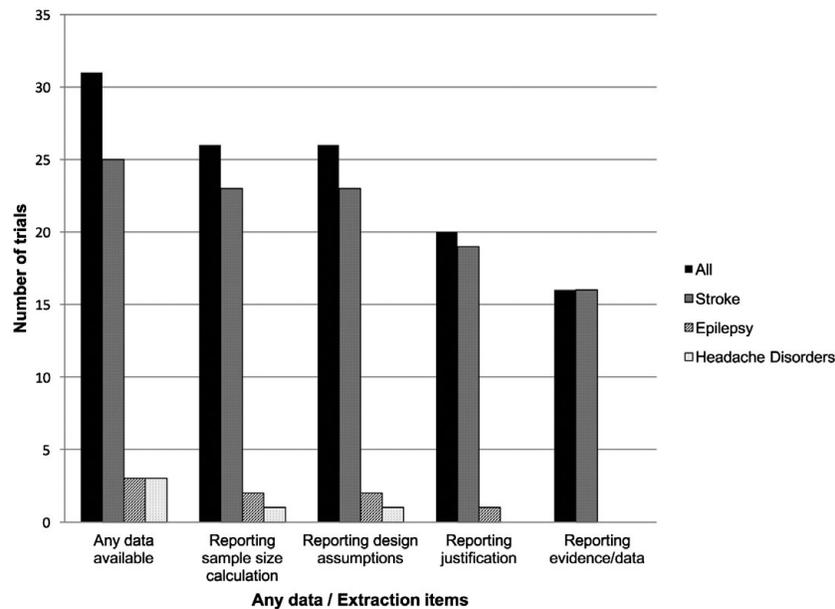


Figure 3. Data availability.

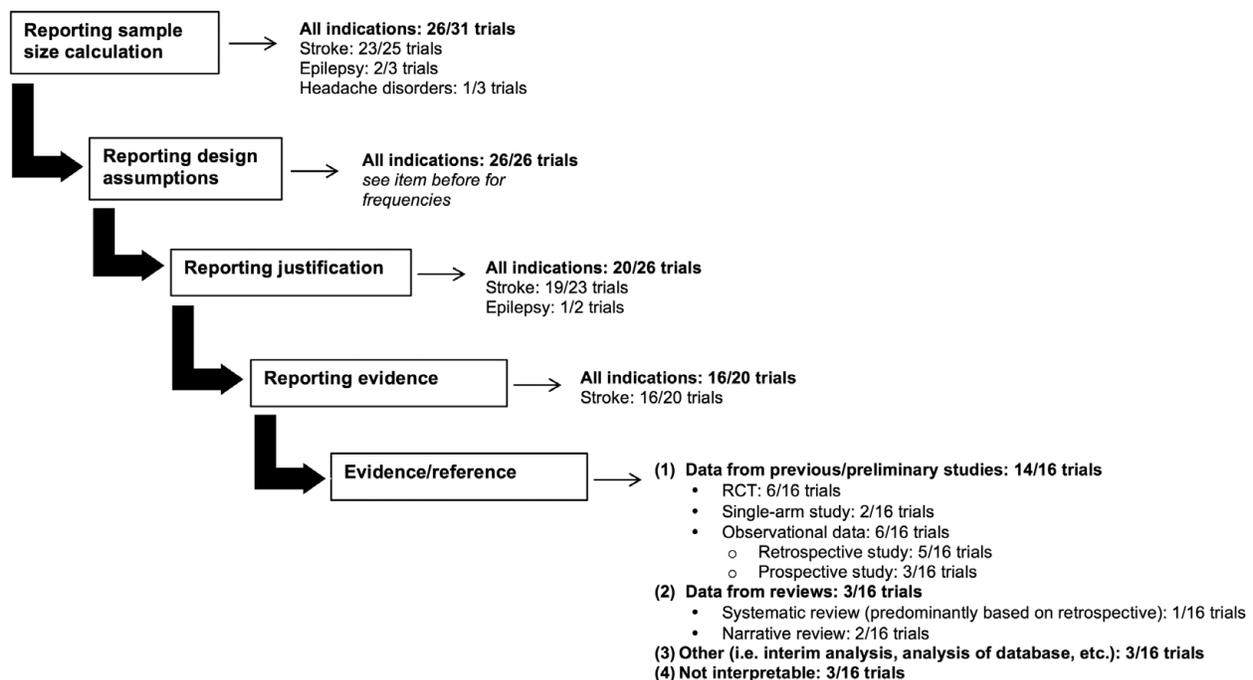


Figure 4. Sample size reporting (multiple sources possible).

Other data used for parameter estimation were all of lower level of evidence, including single-arm studies, narrative reviews, and results from interim analyses. This finding is particularly important given the large influence that parameter estimation has on sample size calculation and consequently on the reliability of the study outcomes. For example, the recent study by Castonguay et al. reviewed the accuracy with which the median progression-free survival (PFS) and overall

survival (OS) in the control arm had been estimated in studies of epithelial ovarian cancer between 2000 and 2010 (21). Their results highlight that PFS and OS of the control arm have frequently been underestimated, which means that the initial statistical assumptions of these trials may have been inaccurate. In those studies in which significant underestimation arose, the authors noted the negative impact on the power to detect absolute differences in the respective endpoints. This decrease

in power consequently compromises the ultimate results of a trial, even before it has started.

Considering that our sample focused on trials regarding high-risk MDs, specifically from a patient safety perspective, the limited transparency documented in this systematic review is alarming. Very recently, Rathi et al. aimed to characterize the clinical evidence generated for high-risk therapeutic devices initially approved by means of the US Food and Drug Administration (FDA) Premarket Approval (PMA) pathway. Their results support our findings by showing that the amount and quality of evidence generated over the total product life cycle of MDs varies widely. Most devices have been, or will be, evaluated by only a few studies, which often focus on surrogate markers in small numbers of patients followed up over short periods of time and study indications that differ from the original FDA-approved indication (22).

By now, the issue of inadequate or missing evidence has found its way into public discourse, as stakeholders have identified it as an obstacle for new technologies trying to enter the market. More precisely, it is postulated that healthcare decision makers choose not to introduce innovative treatment options in such situations and instead wait until stronger evidence has been generated; for example, by implementing the coverage with evidence development approach (23). One project aiming at mitigating this evidence gap is the core protocol pilot within the EUnetHTA network, which attempts to develop and test a common methodological basis for additional evidence generation on a European level (24).

Strengths and Limitations

The main strength of this study is the comprehensive and systematic approach to the identification of RCTs in two trial registries in the area of CNS diseases. To our knowledge, this is the first review regarding reporting quality in studies of MDs, primarily based on original research protocols. However, we acknowledge several limitations.

We chose to focus our search on trial registries to minimize publication bias. However, because trial registration is still not obligatory in Europe, there is a potential risk that our search did not capture some relevant studies.

Our analysis is mainly based on stroke trials, as only scarce data were available for the indications headache disorders and epilepsy. This consequently restricts generalizability of our findings in the area of CNS diseases.

Another limitation is that for four trials, contact information was not available or erroneous, which highlights the need for updating email addresses of responsible parties in the registries. In addition, as the items of the registries do not capture this information, we had to make this assessment ourselves. To ensure a structured and understandable approach of assessment, we used the MD guidance manual published by the European Commission (EC) (19).

Although a duplicate data check was performed aiming at reducing bias, the assessment of the risk classification may still contain errors. However, we believe that our results are valid because we focused on high-risk MDs in general and errors might have occurred only within the high-risk classes IIB and III. Lastly, we did not assess whether the trials in our sample were properly powered or whether the assumptions made were appropriate.

Implications for Policy

Our findings have relevant implications for policy. First, our results show the importance of an improvement in data availability. We recommend that journals enforce inclusion of protocols for all submissions to enable public access to original research data. Although several journals have already made protocol submission with a manuscript as a condition, there is much room for improvement. Furthermore, even if guidelines on reporting exist, more efforts are needed to ensure high quality reporting. Therefore, we suggest tightening submission requirements of publications by a stricter compliance to CONSORT recommendations.

In addition, as ethics committees and regulatory authorities such as the German Federal Institute for Drugs and Medical Devices receive the results prior to scientific review, they have an important gate-keeping role in ensuring that submitted trials are adequately powered and well documented. Moreover, when advising study investigators, they should also address statistical elements such as the realistic calculation of sample size. Given the importance of trial protocols, guidelines for writing study reports that help to facilitate complete documentation of key trial elements, such as SPIRIT, should be routinely used (25). In addition, current initiatives such as the EUnetHTA pilot core project help to give clear guidance on complete transparent and standardized reporting of key methodological parameters in RCTs.

CONCLUSION

Our research confirms previous findings documenting a lack of public access to original data and poor reporting of sample size calculation, especially with respect to underlying evidence supporting study design assumptions. Consequences might be misinterpretations of study results and the inability to make proper decisions in healthcare. Transparent reporting of sample size calculation including justification of underlying assumptions is important from a scientific as well as an ethical point of view to indicate the quality of how a trial has been conducted. Journal editors, ethics committees, and regulatory authorities play a key role in facilitating improved data availability, as well as correct, transparent, and standardized reporting. These factors are necessary to enable critical evaluation of RCTs.

SUPPLEMENTARY MATERIAL

Supplementary Table 1:

<https://doi.org/10.1017/S0266462317000265>

Supplementary Table 2:

<https://doi.org/10.1017/S0266462317000265>

Supplementary Table 3:

<https://doi.org/10.1017/S0266462317000265>

Supplementary Table 4:

<https://doi.org/10.1017/S0266462317000265>

CONFLICTS OF INTEREST

The authors report no conflicts of interest regarding this work.

REFERENCES

- Baeyens H, Poupez C, Slegers P, et al. Towards a guided and phased introduction of high-risk medical devices in Belgium; KCE report 249. 2015. https://kce.fgov.be/sites/default/files/page_documents/KCE249_High-risk%20medical%20devices_Report.pdf (accessed September 17, 2015).
- Persistence Market research. Neurostimulation Devices Market: Global Industry Analysis and Forecast to 2020 [Press Release]. <http://www.mynewsdesk.com/us/persistence-market-research-2/pressreleases/neurostimulation-devices-market-global-industry-analysis-and-forecast-to-2020-1054197> (accessed September 17, 2015).
- Storz P, Kolpatzik K, Perleth M, et al. Future relevance of genetic testing: A systematic horizon scanning analysis. *Int J Technol Assess Health Care*. 2007;23:495-504.
- Mills EJ, Wu P, Gagnier J, et al. The quality of randomized trial reporting in leading medical journals since the revised CONSORT statement. *Contemp Clin Trials*. 2005;26:480-487.
- International Conference on Harmonisation. ICH harmonised tripartite guideline. Statistical principles for clinical trials E9; Current Step 4, version dated 5 February 1998. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step_4/E9_Guideline.pdf (accessed September 17, 2015).
- Weaver CS, Leonardi-Bee J, Bath-Hextall FJ, et al. Sample size calculations in acute stroke trials: A systematic review of their reporting, characteristics, and relationship with outcome. *Stroke*. 2004;35:1216-1224.
- Chan AW, Hrobjartsson A, Jorgensen KJ, et al. Discrepancies in sample size calculations and data analyses reported in randomised trials: Comparison of publications with protocols. *BMJ*. 2008;337:a2299.
- Revez L, Cortes-Jofre M, Asenjo Lobos C, et al. Influence of trial registration on reporting quality of randomized trials: Study from highest ranked journals. *J Clin Epidemiol*. 2010;63:1216-1222.
- Wieseler B, Wolfram N, McGauran N, et al. Completeness of reporting of patient-relevant clinical trial outcomes: Comparison of unpublished clinical study reports with publicly available data. *PLoS Med*. 2013;10:e1001526.
- Julious S. *Sample sizes for clinical trials*. Boca Raton: Chapman and Hall/ CRC; 2009.
- Campbell M, Machin D, Walters S. *Medical statistics. A textbook of health science*. 4th ed. Chichester: John Wiley & Sons; 2007.
- Boutron I, Moher D, Altman DG, et al. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: Explanation and elaboration. *Ann Intern Med*. 2008;148:295-309.
- Jones G, Abbasi K. Trial protocols at the BMJ. *BMJ* 2004;329:1360.
- McNamee D, James A, Kleinert S. Protocol review at The Lancet. *Lancet*. 2008;372:189-190.
- Charles P, Giraudeau B, Dechartres A, et al. Reporting of sample size calculation in randomised controlled trials: Review. *BMJ*. 2009;338:b1732.
- Clark T, Berger U, Mansmann U. Sample size determinations in original research protocols for randomised clinical trials submitted to UK research ethics committees: Review. *BMJ*. 2013;346:f1135.
- Rutterford C, Taljaard M, Dixon S, et al. Reporting and methodological quality of sample size calculations in cluster randomized trials could be improved: A review. *J Clin Epidemiol*. 2015;68:716-723.
- Toerien M, Brookes ST, Metcalfe C, et al. A review of reporting of participant recruitment and retention in RCTs in six major journals. *Trials*. 2009;10:52.
- European Commission, Directorate-General for Health and Food Safety. MEDICAL DEVICES: Guidance document. Classification of medical devices. MEDDEV 2.4/1 Rev.9 .2010. http://ec.europa.eu/health/medical-devices/files/meddev/2_4_1_rev_9_classification_en.pdf (accessed September 17, 2015).
- Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869.
- Castonguay V, Wilson MK, Diaz-Padilla I, et al. Estimation of expectedness: Predictive accuracy of standard therapy outcomes in randomized phase 3 studies in epithelial ovarian cancer. *Cancer*. 2015;121:413-422.
- Rathi VK, Krumholz HM, Masoudi FA, et al. Characteristics of clinical studies conducted over the total product life cycle of high-risk therapeutic medical devices receiving FDA premarket approval in 2010 and 2011. *JAMA*. 2015;314:604-612.
- Olberg B, Perleth M, Busse R. The new regulation to investigate potentially beneficial diagnostic and therapeutic methods in Germany: Up to international standard? *Health Policy*. 2014;117:135-145.
- EUnetHTA. Public consultation on the second draft of: Core protocol Pilot for Additional Evidence Generation (AEG). 2015. <http://www.eunetha.eu/news/closed-public-consultation-second-draft-core-protocol-pilot-additional-evidence-generation-aeg> (accessed September 19, 2015).
- Chan AW, Tetzlaff JM, Gotzsche PC, et al. SPIRIT 2013 explanation and elaboration: Guidance for protocols of clinical trials. *BMJ*. 2013;346:e7586.