

# BMJ Open Validity of sample sizes in publications of randomised controlled trials on the treatment of age-related macular degeneration: cross-sectional evaluation

Sabrina Tulka, Berit Geis, Christine Baulig, Stephanie Knippschild, Frank Krummenauer

**To cite:** Tulka S, Geis B, Baulig C, *et al.* Validity of sample sizes in publications of randomised controlled trials on the treatment of age-related macular degeneration: cross-sectional evaluation. *BMJ Open* 2019;**9**:e030312. doi:10.1136/bmjopen-2019-030312

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-030312>).

Received 11 March 2019  
Revised 22 August 2019  
Accepted 19 September 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

Institute for Medical Biometry and Epidemiology, University Witten Herdecke Faculty of Health, Witten, Germany

## Correspondence to

Sabrina Tulka;  
sabrina.tulka@uni-wh.de

## ABSTRACT

**Objective** The aim of this cross-sectional study was to examine the completeness and accuracy of the reporting of sample size calculations in randomised controlled trial (RCT) publications on the treatment of age-related macular degeneration (AMD).

**Methods** A sample of 97 RCTs published between 2004 and 2014 was reviewed for the calculation of their sample size. It was examined whether a (complete) description of the sample size calculation was presented. Furthermore, the sample size was recalculated, whenever possible based on the published details, in order to verify the reported number of patients.

**Primary outcome measure** The primary endpoint of this cross-sectional investigation was a described sample size calculation that was reproducible, complete and correct (maximum tolerated deviation between reported and replicated sample size  $\pm 2$  participants per trial arm).

**Results** A total of 50 publications (52%) did not provide any information on the justification of the number of patients included. Only 17 publications (18%) provided all the necessary parameters for recalculation; 8 of 97 (8%, 95%-CI: 4% to 16%) publications achieved the primary endpoint. The median relative deviation between reported and recalculated sample sizes was 1%, with a range from  $-43\%$  to  $+66\%$ .

**Conclusion** Although a transparent sample size legitimisation is a crucial determinant of an RCT's methodological validity, more than half of the RCT publications considered failed to report them. Furthermore, reported sample size legitimisations were often incomplete or incorrect. In summary, clinical authors should pay more attention to the transparent reporting of sample size calculation, and clinical journal reviewers may opt to reproduce reported sample size calculations.

**Synopsis** More than half of the analysed RCT publications on the treatment of AMD did not report a transparent sample size calculation. Only 8% reported a complete and correct sample size calculation.

## INTRODUCTION

Each patient study should be based on a valid statistical sample size calculation in order to reveal significant findings under assurance of a sufficiently high statistical power. Sample

## Strengths and limitations of this study

- The validity of sample size calculations for randomised controlled trial (RCT) publications on the treatment of age-related macular degeneration (AMD) has not been investigated so far.
- The data extraction was performed by means of a consensus rating of two biometricians, thereby ensuring outcome validity.
- The AMD results cannot be extrapolated onto other ophthalmological diseases (RCT publications on these); it is not clear how AMD-specific the findings are.
- The reviewers were not blinded towards the journals, publications and authors, so that a reviewer bias could not be excluded.

size calculation is thereby based on statistical as well as clinical assumptions (clinically relevant effects between therapeutic alternatives) for the primary clinical endpoint of a study. A statistical sample size calculation is one of the most crucial determinants of the validity of a trial's result.<sup>1</sup>

As a reporting guideline for publications of randomised controlled trials (RCTs), the Consolidated Standards of Reporting Trials (CONSORT) statement<sup>2</sup> demands a complete justification of the sample sizes. CONSORT requires authors to describe all necessary elements of a sample size calculation to provide a complete and transparent description. This includes the expected effect size characterising the clinically relevant difference between the treatment samples as parameterised by the trial's primary clinical endpoint, as well as the intended levels of significance and power. In strict accordance, the International Conference on Harmonisation (ICH) item 3.5 of the ICH guideline E9 requires a complete description of sample size calculation in the protocol of every clinical

trial (ICH includes guidelines for conducting clinical trials in Europe, the USA and Japan). In addition, a justification for the expected effect size should be reported.<sup>3</sup>

Despite the availability of both RCT reporting standards for longer than two decades, several investigations<sup>4-9</sup> identified clinical trials which either do not provide any information on sample size calculation or incorrect sample sizes in their publication. Bearing these findings in mind, the aim of this study was to examine whether publications of RCTs on age-related macular degeneration (AMD) treatment reported complete and correct sample size calculations: It is expected that RCTs on invasive and drug therapies for severe diseases will be monitored with the highest standard of care. Methodological deficits detected for these RCTs could potentially be even more serious in studies on less invasive therapies. Due to the research focus of ophthalmology, AMD was chosen as an ophthalmological disease, whose studies should fulfil this requirement.

## METHODS

### Search strategy and RCT publication selection

This study was an addition to a project on RCT search strategies. A PubMed search was conducted to identify all eligible RCT publications on AMD healthcare. The search was performed based on the following terms: 'macular degeneration', 'randomised controlled trial' and 'published between: 1/1/2004 and 12/31/2013'. Literature research was limited to the English language (table 1). Two independent parallel reviewers excluded inappropriate articles (CB and SK). Publications not affiliated to AMD, without randomisation, publications with an inappropriate study design and non-English publications were to be excluded from the analysis. Of 673 possible RCTs identified by this search, a total of 133 remained eligible for evaluation; for further description of this RCT publication pool and details on the underlying electronic search strategy, see Baulig *et al.*<sup>10</sup> From this publication pool a series of 97 RCT publications (see sample size calculation) was analysed.

### Data extraction

Each publication and supplementary material (including previous publications, trial registration and

supplementary files when referred to in the publication) was first screened to determine whether information on sample size calculation was provided. This information was extracted from the publication whenever statistical arguments were provided (eg, legitimisation of net sample sizes by referring to budgetary limitations of investigators was not accepted as a methodologically valid sample size calculation). The level of significance, statistical power, the expected effect size and the statistical methods applied for analysis and thereby for sample size calculation were extracted. This process of raw data extraction was performed by means of a consensus rating of two biometricians (ST and FK).

In addition, further editorial information was documented on characteristics of the publications: the year of publication, the underlying journal's Thompson & Reuter impact factor (IF) for the year of publication (ISI Web of Science, table 2), industrial funding, statistical support and the number of trial centres.

### Primary endpoint

This investigation's primary endpoint was achieved by an RCT publication, when a reproducible, complete and correct description of the sample size calculation was reported in that publication, and recalculation/reproduction of the reported sample size was possible with a maximum difference between a reported and replicated sample size of  $\pm 2$  persons per trial arm.

### Reproduction of sample size calculation reports

Replication of the reported sample size calculations was done using the software nQuery Advisor Version 4.0 for Windows. The extracted data on sample size calculation were entered into this programme according to the choice of analysis methods as declared by the respective publications' Statistical methods section. The replicated sample size was then compared with the reported sample size.

If information necessary for recalculation (ie, one of the parameters mentioned earlier) was missing or reported parameters were deemed wrong, the corresponding details were imputed whenever possible. For example, some publications provided explicit information on the underlying significance level but did not explicitly mention whether this significance level was

**Table 1** Full RCT search strategy

Search	Query	Items found	Time
"Search (((macular degeneration) AND ("2004/01/01"(Date - Publication): "2013/12/31"(Date - Publication)))) AND english(Language)) AND "randomized controlled trial"(Publication Type)"	#12	673	07:44:11
"Search (((macular degeneration) AND ("2004/01/01"(Date - Publication): "2013/12/31"(Date - Publication)))) AND english(Language)"	#11	11 737	07:44:00
"Search (macular degeneration) AND ("2004/01/01"(Date - Publication): "2013/12/31"(Date - Publication))"	#10	13 068	07:43:41
"Search macular degeneration"	#1	22 957	07:00:29

**Table 2** Journals IF ranges (derived from the ISI Web of Science) for journals having published the 97 RCT publications used for sample size evaluation, frequency of analysed RCT publications per journal

Journal	Journal-wise number of RCT publications	Journal IF range for years of the RCT publications under consideration
<i>Acta Ophthalmologica</i>	4	1.028 – 1.867
<i>Acta Ophthalmologica Scandinavica</i>	1	0.986
<i>American Journal of Ophthalmology</i>	5	2.332 – 3.631
<i>Archives of Ophthalmology</i>	5	2.926 – 3.274
<i>Atherosclerosis Supplements</i>	1	9.667
<i>Biomedical Papers</i>	1	1.200
<i>BMJ</i>	1	13.471
<i>British Journal of Nutrition</i>	1	2.764
<i>British Journal of Ophthalmology</i>	6	2.725 – 2.934
<i>Clinical Hemorheology and Microcirculation</i>	1	0.000
<i>Clinical Rehabilitation</i>	1	1.118
<i>Current Medical Research and Opinion</i>	1	2.604
<i>European Journal of Ophthalmology</i>	1	0.912
<i>Eye (Lond)</i>	5	1.818 – 1.974
<i>Graefes Archive for Clinical and Experimental Ophthalmology</i>	3	1.498 – 2.333
<i>Investigative Ophthalmology and Visual Science</i>	2	3.661 – 3.766
<i>Journal of Clinical Neuroscience</i>	1	1.165
<i>JAMA Ophthalmology</i>	1	4.488
<i>New England Journal of Medicine</i>	4	37.841 – 52.414
<i>Nutrients</i>	1	3.148
<i>Nutrition</i>	1	3.046
<i>Ophthalmologica</i>	1	0.986
<i>Ophthalmology</i>	34	3.210 – 6.170
<i>Optometry</i>	3	0.000 – 0.741
<i>PLoS One</i>	1	3.534
<i>Proceedings of the National Academy of Sciences of the USA</i>	1	9.681
<i>Retina</i>	8	2.774 – 3.177
<i>Scandinavian Journal of Occupational Therapy</i>	1	0.000
<i>Transactions of the American Ophthalmological Society</i>	1	0.000

IF ranges denote changes in journal IF over several RCT publication years.  
IF, impact factor; RCT, randomised controlled trial.

corrected for multiplicity in the sample size calculation for a multiple trial arm comparison (eg, by means of a Bonferroni correction); in such cases, the recalculation assumed the methodologically correct approach with regard to the study design at hand; that is, in general, the sample size recalculation had to match the study design, even if the published sample size calculation did not.

Sample size calculation reports omitting details on the following design parameters were not classified as incomplete, whenever the actual methods choice for analysis and planning could be assumed by means of available context information: two-tailed test (superiority), one-tailed test

(non-inferiority), statistical test (if explained elsewhere in the Methods section or the Results section), technical continuity correction details (eg, for the  $\chi^2$  test), hierarchical interdependence of multiple primary endpoints and hypothesis.

For one or more of the publications examined, the following parameters had to be imputed based on context information: expected difference (for non-inferiority trials, always assumed '0'), expected standard deviation (two possibilities: either the value from previous studies mentioned in the publication at hand or backward calculation based on the reported sample size), expected effect

size (two possibilities: either the effect size from another study reported in the publication at hand was imputed or a backward calculation was performed based on the sample size reported in the RCT publication at hand).

### Statistical analysis and sample size calculation

In order to detect an expected frequency of 50% primary endpoint violations - and thereby invalid or non-transparent information on the sample size in at least every second RCT publication - and assuming a confidence level of 95% and  $\pm 10\%$  as the maximum width of the confidence interval (CI) for this expected primary endpoint frequency, a total of 97 publications had to be included in the evaluation.

Statistical analysis of the primary endpoint was then performed by estimating its cross-sectional prevalence by means of the 95% Clopper-Pearson CI. Furthermore, the relative deviation (%) of the reported and recalculated sample sizes was calculated via:

$$\frac{\text{sample size (replicated)} - \text{sample size (reported)}}{\text{sample size (replicated)}} \times 100.$$

To describe the distribution of these studywise differences, medians, quartiles and ranges were estimated; non-parametric boxplots were used as a graphical presentation.

### Patient and public involvement

As this investigation was based on published aggregate data (ie, secondary data evaluation) only, no individual patient contact or individual patient data were involved. In particular, no information from or to patients had to be communicated.

## RESULTS

### RCT publication characteristics

This cross-sectional evaluation comprised 97 RCT publications from 29 journals, of which 30 (31%) were published in a journal with an IF of  $\geq 5$  at the year of publication and 67 (69%) in a journal with an IF of  $\leq 5$ . Fifty-three per cent of the published RCTs were multicentre trials, 51% stated industrial funding and 54% claimed the participation of a statistician or a statistical methods unit. In 83 of 97 (86%) RCT publications, a primary efficacy or effectiveness endpoint was examined.

A total of 50 out of 97 RCT publications did not report any information on sample size calculation (95%-CI: 42% to 62%). Eight descriptions of sample size calculation (8%, 95%-CI: 4% to 16%) were complete and reproducible, so that the underlying RCT publication achieved this investigation's primary endpoint.

The replication of reported sample size calculations was possible for 36 RCT publications (77% of the 47 publications with reported sample size legitimation, 37% of all 97 publications analysed).

Only 17 (18% of 97) publications provided all necessary information to replicate the described sample size calculation, whereas 19 reports were incomplete or incorrect

**Table 3** Frequencies of missing or wrong values in publications with reported sample size calculation

Publications with recalculation via imputed values	
Wrong values	
Power	5
Level of significance	3
Effect size	2
Method (statistical test)	1
Missing values	
Power	–
Level of significance	2
Effect size	12
Publications without recalculation	
Pilot study (one with budgetary limitations)	2
No sample size reported	1
Incorrect or inconclusive planning (power analysis)	8

(table 3) (however, they provided sufficient information to recalculate the sample size using values assumed from the context).

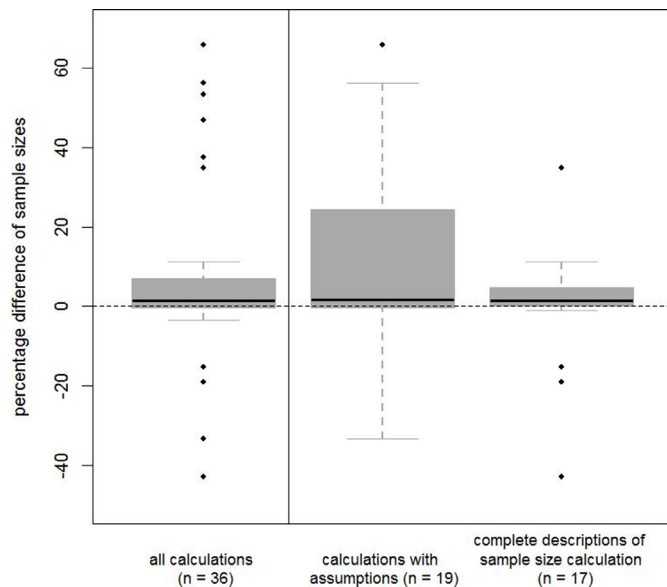
The median percentage difference between the replicated and reported sample sizes was estimated 1% (IQR:  $-1\%$  to  $+5\%$ ), and the median absolute difference between the replicated and reported sample sizes was 1.50 (IQR:  $-1$  to 5.25, range  $-24$  to  $+502$ ) for the 36 publication enabling for recalculations with or without additional assumptions due to incomplete or incorrect input data (figure 1). Maximum deviations were  $-43\%$  (reported  $n=10$ , replicated  $n=7$ ) and  $66\%$  (reported  $n=261$ , replicated  $n=763$ ).

Among those publications reporting complete and correct input data (and thereby not requiring imputation or assumption of parameters,  $n=17$  publications), the median percentage difference between the reported and replicated sample sizes was again estimated 1% (IQR:  $0\%$ – $5\%$ ) with minimum and maximum deviations of  $-43\%$  (reported  $n=10$ , replicated  $n=7$ ) and  $+35\%$  (reported  $n=300$ , replicated  $n=461$ ).

Publications in journals with an IF of  $\leq 5$  in the respective year of publication showed a median percentage difference of 2% (IQR:  $0\%$  to  $+6\%$ ), while sample size calculations in journals with an IF of  $> 5$  showed a median percentage difference of 0% (IQR:  $-1\%$  to  $+3\%$ ). The median percentage difference between RCTs published before 2010 (IQR  $-1\%$  to  $+3\%$ , range:  $-43\%$  to  $66\%$ ) and in 2010 or later (IQR:  $-1\%$  to  $+3\%$ , range:  $-33\%$  to  $56\%$ ) was 1%.

## DISCUSSION

This cross-sectional investigation demonstrated a notable lack of methodological transparency and correctness of sample size calculations in AMD RCT publications



**Figure 1** Boxplots for the relative deviation (%) of reported and recalculated sample size calculations (based on 36 RCT publications providing sufficiently detailed information for a sample size recalculation), presented for all 36 publications as well as stratified for publications with complete information for recalculation (17 RCT publications), and for publications only reporting incomplete or incorrect information and thereby requiring assumptions or corrections for the recalculation of sample sizes (19 RCT publications). Horizontal lines indicate medians and quartiles; vertical lines indicate total ranges to minimum and maximum deviations; diamonds indicate outlier deviations with at least double IQR deviations from the median. RCT, randomised controlled trial.

(and supplementary material or previous publications if referred to in the publication). Only 8% of the 97 RCT publications on the treatment of AMD reported a sample size calculation that was both complete and matched the reported sample size (maximal discrepancy of  $\pm 2$  persons per study group allowed according to inevitable differences due to numerical algorithms applied in calculation software packages).

The reasons for the observed lack in reporting and/or trial implementation quality may vary: for example, one publication described budgetary limitations as an explanation for the enrolled number of patients instead of a statistical rationale. However, more than half of the analysed publications did not report any information on how the included number of patients was calculated (no sample size calculation or other reason). It seems possible that the description of the sample size calculation was deleted, although initially contained, from a publication draft in order to reduce the number of words and thereby adhere to word count limitations (such as required by most clinical journals).

Whatever may have led to the observed deficits in reporting quality cannot be excused by the possible origins hypothesised earlier: the transparent reporting of a sample size calculation is an important tool for assessing whether a study was planned carefully and had

the opportunity of finding significant results in the first place. Moreover, the overall credibility of a study is called into question if a sample size calculation is not reported, making the presumption possible, that the trial never underwent a proper planning phase. Without doubt a transparent sample size justification is necessary to avoid misinterpretation of study results. In summary, there is potential to improve reporting on sample size calculations in publications on AMD treatment. A logistic regression did not reveal factors (IF, funding and year of publication) clearly associated with a study's chance of reaching our study's primary endpoint).

Literature, however, demonstrates that this tendency is by far not AMD specific. The findings of this investigation are in line with the results of other studies that have examined the quality and accuracy of the descriptions of sample size calculations in publications.<sup>4-9</sup> One study analysed sample size calculations in publications, which had appeared in six high impact between 2005 and 2006: a total of 95% of all publications analysed in this study provided information on the calculation of the sample size, whereas of these, 43% did not report all necessary information.<sup>4</sup> Recalculation led to a range of differences between reported and replicated sample sizes from -50% to 50%.

Lee and Tse<sup>5</sup> examined the quality of sample size calculation in 451 RCT publications (published in December 2014 and indexed in PubMed): in 58.1% of the publications, a sample size calculation was described (with recalculation having been possible for 40% of these publications). The comparison of the replicated and reported sample sizes showed a median deviation of 0% (IQR: -4.6% to +3%). Moreover, only 39.7% (25 out of 63) of the sample sizes were identical to the sample sizes stated in trial registers (difference: median: 0%, IQR: -8.1% to +15.1%). A multiple linear regression showed that journals recommending the CONSORT statement and having an IF published articles with more details and smaller deviations between reported and recalculated sample sizes.

In other reviews, 78% (66% complete)<sup>6</sup> and 91.7% (80.3% complete)<sup>7</sup> of anaesthesia publications reported sample size calculations. In RCT publications from the field of dentistry and orthodontics, descriptions of sample size calculations were found in only 29.3%<sup>8</sup> and 29.5%,<sup>9</sup> respectively. The respective differences between the reported and replicated sample sizes were then found to range from -237.5% to 84.2%<sup>8</sup> and -93.3% to 60.6%.<sup>9</sup> Furthermore, there was also a discrepancy between the planned and the actually recruited number of patients (recruited sample size smaller than planned sample size: 23.6%, recruited sample size larger than planned sample size: 58.4%).<sup>11</sup>

Some authors could demonstrate that a later year of publication had a positive effect on the completeness of sample size data.<sup>6 8 12</sup>

Missing sample size calculations were also found in protocols of clinical trials. From 446 protocols, only 42%



reported all necessary elements of a sample size calculation. The replicated sample sizes were identical to the reported sample sizes in only 30% of the trials.<sup>13</sup> In addition, it could be shown that there were also discrepancies between sample sizes in publications and protocols.<sup>14</sup> Another study documented that only 31 out of 71 studies (protocol/publication) provided information on how the sample size was calculated (26 complete descriptions).<sup>15</sup>

### Study limitations

Evaluations and replications of sample size calculations were carried out by one consultant (ST) only (no independent parallel evaluation); however, all replications were discussed with and reviewed by a second consultant (BG), and a consensus was found by the additional review of an experienced and certified biometrician (FK) whenever deemed necessary or appropriate. A further limitation is that the assessment was not performed as a blinded procedure; that is, the reviewers were not blinded towards the journals, publications and authors, possibly having resulted in a reviewer bias (eg, in rating a value as wrong). Note in addition, that several RCT author teams contributed more than one RCT report to the 97 publications' pool, yet implying potential increase of the effect of such bias mechanisms. In addition, only a limited period of time (2004–2014) was examined. It can be assumed that publications published after 2014 may have a higher frequency of describing sample size calculation as the journals increasingly recommend the strict use of reporting standards such as those comprised in the CONSORT statement. This period of time was chosen as this project was an add-on to a project on search strategies.<sup>16</sup> A follow-up project on publications after 2014 is planned.

From the pragmatic clinical trial investigator's perspective, this investigation's primary endpoint may furthermore have been designed overly strict for publications on RCTs on larger patient samples, as only a discrepancy of  $\pm 2$  subjects was allowed from the numerical implementation perspective. Reanalyses based on a secondary endpoint allowing for a maximum discrepancy between recalculated and reported sample sizes of  $\pm 10\%$ , however, demonstrated a similar overall tendency as observed for the primary endpoint: 12% of 97 of the publications had a sample size calculation that reached this secondary endpoint.

Considering the validity of reported sample size calculations, however, naturally calls for the reassessment of the 'own' sample size legitimisation, yet actually having been based on the 'incorrect' assumption of about 50% invalid descriptions of sample size calculation contrasted to the observed prevalence of 92%. For the CI of the observed prevalence of 8% correct and complete sample size justifications, the recruited number of 97 publications must be admitted as having been chosen too small: the 95% confidence estimation of such an expected frequency would rather be based on requiring a maximum CI length of, say,  $\pm 2\%$  instead of  $\pm 10\%$  (as required for the 50% prevalence

assumption; see previous discussion). As a consequence, a total of 707 RCT publications would have been necessary for evaluation, yet demonstrating the essential 'drawback' of sample size calculation—you only know, whether the underlying assumption and thereby the result of sample size calculations were correct after you have performed the trial. From this perspective, some of the 97 RCT publications might have omitted a sample size calculation report just for this simple reason—the initial sample size assumptions were substantially wrong. Nevertheless, transparent reporting still would encourage the reporting of the underlying assumptions and thereby explain the difference between expected and observed outcome, as well as required and achieved statistical power.

### CONCLUSION

Although the CONSORT statement is available since 1996, more than half of the publications analysed here did not report a sample size calculation. Described sample size calculations were often incorrect (calculation and practically applied sample size did not match) or incomplete (not all necessary elements were reported). This demonstrates the substantial need for improvement and, at the same time, provides constructive lines for implementation of the latter: for example, each journal could provide explicit instructions and example-illustrated guidelines for the reporting of sample size calculations. Furthermore, qualified statisticians should be involved in the planning process of a study design by means of correct sample calculations, and their active involvement in the publication process should be invoked by journals, for example, by requiring written confirmation of explicit contributions to the Methods section of a submitted article. As a consequence, statisticians will be assisted in insisting that their calculation rationale is included in any resulting publication.

Editors and reviewers should also require each author team to provide detailed information on sample size calculations to ensure its reproducibility, at least by means of electronic supplements; the expert review of clinical articles on RCTs could, in addition, mandatorily involve qualified statisticians who could be encouraged to explicitly recalculate reported sample sizes regarding their crucial impact on the overall trial result interpretation.

**Acknowledgements** The authors thank the Leonard Stinnes Foundation, which made this research work possible in terms of the financial means for ST's research grant and, furthermore, Ms Tara Rödter, MD, for a native speaker revision of the first draft of the manuscript.

**Contributors** ST extracted the randomised controlled trial publications' relevant outcome data for the sample size calculations (parallel independent evaluation), performed the sample size recalculations and the statistical analysis of the recalculation data, and wrote this systematic review's first draft. BG double-checked the outcome data extraction and assisted in the sample sizes' evaluation and recalculation; furthermore she reviewed the manuscript draft. CB carried out the publication search, excluded inappropriate articles and revised the manuscript. SK carried out the publication search, excluded inappropriate articles and revised the manuscript. FK wrote the grant application to the Leonhard Stinnes Foundation, designed this investigation, extracted relevant outcome data (parallel independent

evaluation) for selected RCT publications, assisted in sample size recalculations, thoroughly revised the first draft of the manuscript and contributed major parts to the second draft version.

**Funding** This work was supported by Leonard Stinnes Foundation grant number, internal reference, KS 11535.

**Competing interests** This systematic review was conducted by ST (MSc Statistics) within the framework of a full-time placement as a research assistant. This placement was funded by a 24-month research grant received from the Leonard Stinnes Foundation (internal reference: KS 11535). The study has no political conflict of interest, neither in terms of content nor with regard to the results. The results presented in this manuscript are part of the doctoral thesis of ST to be submitted to the Faculty of Health of Witten/Herdecke University to achieve the doctoral degree 'Dr rer medic' (Doctor of Theoretical Medicine). Furthermore, the results contained in this article have already been presented by means of an oral presentation at the annual meeting of the German Region of the International Biometric Society (Frankfurt/Main, Germany, March 2018) and by means of a poster presentation at the annual meeting of the German Ophthalmic Surgeons (Nuremberg, Germany, June 2018), where the presentation was awarded with the 2018 poster prize of the Ophthalmic Surgeons.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

- Krummenauer F, Dick B, Schwenn O, *et al*. The determination of sample size in controlled clinical trials in ophthalmology. *Br J Ophthalmol* 2002;86:946–7.
- 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.
- ICH. E9 statistical principles for clinical trials. London, UK International Conference on Harmonisation; 1998.
- Charles P, Giraudeau B, Dechartres A, *et al*. Reporting of sample size calculation in randomised controlled trials: review. *BMJ* 2009;338.
- Lee PH, Tse ACY. The quality of the reported sample size calculations in randomized controlled trials indexed in PubMed. *Eur J Intern Med* 2017;40:16–21.
- Chow JTY, Turkstra TP, Yim E, *et al*. Sample size calculations for randomized clinical trials published in anesthesiology journals: a comparison of 2010 versus 2016. *Can J Anaesth* 2018;65:611–8.
- Abdulatif M, Mukhtar A, Obayah G. Pitfalls in reporting sample size calculation in randomized controlled trials published in leading anaesthesia journals: a systematic review. *Br J Anaesth* 2015;115:699–707.
- Koletsis D, Fleming PS, Seehra J, *et al*. Are sample sizes clear and justified in RCTs published in dental journals? *PLoS One* 2014;9:e85949.
- Koletsis D, Pandis N, Fleming PS. Sample size in orthodontic randomized controlled trials: are numbers justified? *Eur J Orthod* 2014;36:67–73.
- Baulig C, Krummenauer F, Geis B, *et al*. Reporting quality of randomised controlled trial Abstracts on age-related macular degeneration health care: a cross-sectional quantification of the adherence to consort Abstract reporting recommendations. *BMJ Open* 2018;8:e021912.
- Castellini G, Gianola S, Bonovas S, *et al*. Improving power and sample size calculation in rehabilitation trial reports: a methodological assessment. *Arch Phys Med Rehabil* 2016;97:1195–201.
- Martin J, Taljaard M, Girling A, *et al*. Systematic review finds major deficiencies in sample size methodology and reporting for stepped-wedge cluster randomised trials. *BMJ Open* 2016;6:e010166.
- 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.
- Chan A-W, Hróbjartsson A, Jørgensen KJ, *et al*. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ* 2008;337.
- Olberg B, Perleth M, Felgentraeger K, *et al*. Quality of sample size estimation in trials of medical devices: high-risk devices for neurological conditions as example. *Int J Technol Assess Health Care* 2017;33:103–10.
- Baulig C, Niemeyer L, Krummenauer F, *et al*. Comparison of review search strategies “hand search” versus “electronic literature research in PubMed und EMBASE®”: Relative precision and implication for the quantitative result of a meta-analysis, illustrated by the meta-estimate of the expectable dropout rate in RCTs for the treatment of AMD patients, under review, 2019.