

# BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com)

# BMJ Open

## Validity of sample sizes in RCT publications on the treatment of AMD: a cross-sectional evaluation

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2019-030312
Article Type:	Research
Date Submitted by the Author:	11-Mar-2019
Complete List of Authors:	Tulka, Sabrina; University Witten Herdecke Faculty of Health, Institute for Medical Biometry and Epidemiology, Geis, Berit; University Witten Herdecke Faculty of Health, Institute for Medical Biometry and Epidemiology Krummenauer, Frank; University Witten Herdecke Faculty of Health, Institute for Medical Biometry and Epidemiology,
Keywords:	sample size calculation, RCT publication, transparent reporting, recalculation

SCHOLARONE™  
Manuscripts

# Validity of sample sizes in RCT publications on the treatment of AMD: a cross-sectional evaluation

Sabrina Tulka, Berit Geis, Frank Krummenauer

University Witten / Herdecke Faculty of Health, Institute for Medical Biometry and Epidemiology, 58448 Witten, Germany

Correspondence:

Sabrina Tulka, M.Sc.  
University Witten / Herdecke Faculty of Health,  
Institute for Medical Biometry and Epidemiology,  
Alfred Herrhausen-Straße 50, D-58448 Witten (Germany)  
phone: ++ 49 2302 926 471  
fax: ++ 49 2302 926 44785  
e-mail: sabrina.tulka@uni-wh.de

**Key Words:** sample size calculation, RCT publication, transparent reporting, recalculation

**Synopsis:** More than half of the analyzed 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.

(word count: 2981)

## Abstract

1  
2  
3 **Objective:** The aim of this cross-sectional study was to examine the completeness  
4 and accuracy of the reporting of sample size calculations in RCT publications on the  
5 treatment of age-related macular degeneration (AMD).  
6  
7

8 **Methods:** A sample of 97 randomized controlled trials published between 2004 and  
9 2014 was reviewed for the calculation of their sample size. It was examined whether  
10 a (complete) description of the sample size calculation was presented. Furthermore,  
11 the sample size was recalculated, whenever possible based on the published details,  
12 in order to verify the reported number of patients.  
13  
14

15 **Primary outcome measure:** The primary endpoint of this cross-sectional  
16 investigation was a described sample size calculation that was reproducible,  
17 complete and correct (maximum tolerated deviation between reported and replicated  
18 sample size  $\pm 2$  participants per trial arm).  
19  
20

21 **Results:** A total of 50 publications (52%) did not provide any information on the  
22 justification of the number of patients included. Only 17 publications (18%) provided  
23 all the necessary parameters for recalculation, 8 of 97 (8%, 95%-CI: [4%; 16%])  
24 publications achieved the primary endpoint. The median relative deviation between  
25 reported and recalculated sample size was 1% with a range from -43% to +66%.  
26  
27

28 **Conclusion:** Although a transparent sample size legitimization is a crucial determinant  
29 of an RCT's methodological validity, more than half of the RCT publications  
30 considered failed to report them. Furthermore, reported sample size legitimations  
31 were often incomplete or incorrect. In summary, clinical authors should pay more  
32 attention to the transparent reporting of sample size calculation and clinical journal  
33 reviewers may opt to reproduce reported sample size calculations.  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

44 (255 words)  
45  
46  
47  
48

#### 49 **Strengths and Limitations of this study:**

- 51 • Validity of sample size calculations for RCT publications on the treatment of  
52 age-related macular degeneration has not been investigated so far.
- 53 • The data extraction was performed by means of a consensus rating of two  
54 biometricians, thereby ensuring outcome validity.
- 55 • The AMD results cannot be extrapolated onto other ophthalmological diseases  
56 (RCT publications on these); it is not clear, how AMD-specific the findings are.  
57  
58  
59  
60

- The reviewers were not blinded towards the journals, publications and authors, so that a reviewer bias cannot be excluded.

## 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 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[1].

As a reporting guideline for publications of randomized controlled trials (RCT) the CONSORT statement[2] 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 characterizing the clinically relevant difference between the treatment samples as parameterized by the trial's primary clinical endpoint, as well as the intended levels of significance and power. In strict accordance, the 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 denotes the „International Conference on Harmonisation“ guidelines for conducting clinical trials in Europe, the United States of America and Japan). In addition, a justification for the expected effect size should be reported[3].

Despite the availability of both RCT reporting standards for longer than two decades, several investigations[4]-[9] 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 AMD treatment reported complete and correct sample size calculations: RCTs on the medical care of patients suffering from AMD can be expected to be implemented at maximum standards of methodological quality as most of these studies considered medicinal drugs and were thereby thoroughly monitored by regulatory authorities. A lack of methodological quality in the publication of such RCTs – as indicated by non-transparent or incomplete sample size

1  
2  
3 justifications – would thus indicate these deficits are to be expected in other fields of  
4 medicine in an ever more extreme order.  
5  
6  
7

## 8 **METHODS**

### 9 **Primary endpoint, statistical analysis and sample size calculation**

10 This investigation's primary endpoint was achieved by an RCT publication, when a  
11 reproducible, complete and correct description of the sample size calculation was  
12 reported in that publication, and recalculation / reproduction of the reported sample  
13 size was possible with a maximum difference between reported and replicated  
14 sample size of  $\pm 2$  persons per trial arm.  
15  
16  
17  
18  
19

20 In order to detect an expected frequency of 50% primary endpoint violations – and  
21 thereby invalid or non-transparent information on the sample size in at least every  
22 second RCT publication – and assuming a confidence level of 95% and  $\pm 10\%$  as the  
23 maximum width of the confidence interval for this expected primary endpoint  
24 frequency, a total of 97 publications had to be included in the evaluation.  
25  
26  
27  
28

29 Statistical analysis of the primary endpoint was then performed by estimating its  
30 cross-sectional prevalence by means of the 95% Clopper-Pearson confidence  
31 interval. Furthermore the relative deviation [%] of the reported and recalculated  
32 sample sizes was calculated via:  
33  
34  
35

$$36 \frac{\text{sample size (replicated)} - \text{sample size (reported)}}{\text{sample size (replicated)}} * 100.$$

37  
38

39 To describe the distribution of these study-wise differences', medians, quartiles and  
40 ranges were estimated; non-parametric boxplots were used as a graphical  
41 presentation.  
42  
43  
44  
45

### 46 **Search strategy and RCT publication selection**

47 A PubMed search was conducted to identify all eligible RCT publications on AMD  
48 health care. The search was performed based on the following terms: "macular  
49 degeneration", "randomized controlled trial" and "published between: 1/1/2004 and  
50 12/31/2013". Literature research was limited to the English language. Two  
51 independent parallel reviewers excluded inappropriate articles (see  
52 acknowledgement). Publications not affiliated to AMD, without randomization,  
53 publications with an inappropriate study design and non-English publications were to  
54 be excluded from the analysis. Of 673 possible RCTs identified by this search a total  
55  
56  
57  
58  
59  
60

1  
2  
3 of 133 remained eligible for evaluation; for further description of this RCT-publication-  
4 pool and details on the underlying electronic search strategy see Baulig et al.[10].  
5 From this publication pool a total of 97 RCT-publications (see sample size calculation  
6 above) were drawn by random (Figure 1).  
7  
8  
9

## 10 11 **Patient and Public Involvement**

12 As this investigation was based on published aggregate data (i.e. secondary data  
13 evaluation) only, no individual patient contact or individual patient data were involved.  
14 In particular, no information from or to patients had to be communicated.  
15  
16  
17  
18  
19

## 20 21 **RCT publication characteristics**

22 This cross-sectional evaluation comprised 97 RCT publications from 29 journals, of  
23 which 30 (31%) were published in a journal with impact factor (IF)  $\geq 5$  at the year of  
24 publication and 67 (69%) in a journal with impact factor  $\leq 5$ . 53% of the published  
25 RCTs were multicenter trials, 51% stated industrial funding and 54% claimed the  
26 participation of a statistician or a statistical methods unit. In 83 of 97 (86%) RCT  
27 publications, a primary efficacy or effectiveness endpoint was examined.  
28  
29  
30  
31  
32  
33

## 34 35 **Data extraction**

36 Each publication and supplementary material was first screened to determine  
37 whether information on sample size calculation was provided. This information was  
38 extracted from the publication whenever statistical arguments were provided (e.g.  
39 legitimization of net sample sizes by referring to budgetary limitations of investigators  
40 was not accepted as a methodologically valid sample size calculation). The level of  
41 significance, statistical power, the expected effect size and the statistical methods  
42 applied for analysis and thereby for sample size calculation were extracted. This  
43 process of raw data extraction was performed by means of a consensus rating of two  
44 biometricians.  
45  
46  
47  
48  
49  
50

51 In addition, further editorial information was documented on characteristics of the  
52 publications: the year of publication, the underlying journal's Thompson & Reuter  
53 impact factor (IF) for the year of publication (ISI Web of Science, Table 1), industrial  
54 funding, statistical support and the number of trial centers.  
55  
56  
57  
58

59 **Table 1:** Journals impact factor ranges (derived from the ISI Web of Science®) for journals  
60 having published the 97 RCT publications used for sample size evaluation, frequency of

analyzed RCT publications per journal; impact factor ranges denote changes in journal impact factor over several RCT publication years.

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

### 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 program 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 (i.e. one of the parameters mentioned above) 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 corrected for multiplicity in the sample size



1  
2  
3 calculation for a multiple trial arm comparison (e.g. by means of a Bonferroni  
4 correction); in such cases, the recalculation assumed the methodologically correct  
5 approach with regard to the study design at hand; i.e. in general the sample size  
6 recalculation had to match the study design, even if the published sample size  
7 calculation did not.  
8  
9

10  
11 Sample size calculation reports omitting details on the following design parameters  
12 were not classified as incomplete, whenever the actual methods choice for analysis  
13 and planning could be assumed by means of available context information: two-tailed  
14 test (superiority), one-tailed-test (non-inferiority), statistical test (if explained  
15 elsewhere in the Methods section or the Results section), technical continuity  
16 correction details (e.g. for the  $X^2$ - Test), hierarchical interdependence of multiple  
17 primary endpoints and hypothesis.  
18  
19

20  
21 For one or more of the publications examined, the following parameters had to be  
22 imputed based on context information: expected difference (for non-inferiority trials,  
23 always assumed "0"), expected standard deviation (two possibilities: either the value  
24 from previous studies mentioned in the publication at hand, or backward calculation  
25 based on the reported sample size), expected effect size (two possibilities: either the  
26 effect size from another study reported in the publication at hand was imputed, or a  
27 backward calculation was performed based on the sample size reported in the RCT-  
28 publication at hand).  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

## 40 **RESULTS**

41 A total of 50 out of 97 RCT publications did not report any information on sample size  
42 calculation (95%-CI: [42%; 62%]). Eight descriptions of sample size calculation (8%;  
43 95%-CI: [4%; 16%]) were complete and reproducible, so that the underlying RCT  
44 publication achieved this investigation's primary endpoint.  
45  
46

47 The replication of reported sample size calculations was possible for 36 RCT  
48 publications (77% of the 50 publications with reported sample size legitimation; 37%  
49 of all 97 publications analyzed).  
50  
51

52 Only 17 (18% of 97) publications provided all necessary information to replicate the  
53 described sample size calculation, whereas 19 reports were incomplete or incorrect  
54 (however, they provided sufficient information to recalculate the sample size using  
55 values assumed from the context).  
56  
57  
58  
59  
60

1  
2  
3 The median percentage difference between the reported and replicated sample sizes  
4 was estimated 1% (interquartile range (IQR): -1% to +5%) for the 36 publication  
5 enabling for recalculations with or without additional assumptions due to incomplete  
6 or incorrect input data (Figure 2). Maximum deviations were -43% (reported n = 10,  
7 replicated n = 7) and 66% (reported n = 261, replicated n = 763).

8  
9  
10  
11 Among those publications reporting complete and correct input data (and thereby not  
12 requiring imputation or assumption of parameters) the median percentage difference  
13 between the reported and replicated sample sizes was again estimated 1%  
14 (interquartile range: 0% to 5%) with minimum and maximum deviations of -43%  
15 (reported n = 10, replicated n = 7) and +35% (reported n = 300, replicated n = 461).

16  
17 Publications in journals with an impact factor of  $\leq 5$  in the respective year of  
18 publication showed a median percentage difference of 2% (IQR: 0% - +6%), while  
19 sample size calculations in journals with an impact factor of  $> 5$  showed a median  
20 percentage difference of 0% (IQR: -1% - +3%). The median percentage difference  
21 between RCTs published before 2010 (IQR: -1% to +3%, range: -43% to 66%) and in  
22 2010 or later (IQR: -1% to +3%, range: -33% to 56%) was 1%.

## 31 32 **DISCUSSION**

33  
34 This cross-sectional investigation demonstrated a notable lack of methodological  
35 transparency and correctness of sample size calculations in AMD RCT-publications.  
36 Only 8% of the 97 RCT-publications on the treatment of AMD reported a sample size  
37 calculation that was both complete and matched the reported sample size (maximal  
38 discrepancy  $\pm 2$  persons per study group allowed according to inevitable differences  
39 due to numerical algorithms applied in calculation software packages).

40  
41 The reasons for the observed lack in reporting and / or trial implementation quality  
42 may vary: For example, some publications described budgetary limitations as an  
43 explanation for the enrolled number of patients instead of a statistical rationale.  
44 However, more than half of the analyzed publications did not report any information  
45 on how the included number of patients was calculated. It seems possible that the  
46 description of the sample size calculation was deleted, although initially contained,  
47 from a publication draft in order to reduce the number of words and thereby adhere to  
48 word count limitations (such as required by most clinical journals).

49  
50 Whatever may have led to the observed deficits in reporting quality cannot be  
51 excused by the possible origins hypothesized above: the transparent reporting of a  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 sample size calculation is an important tool for assessing whether a study was  
4 planned carefully and had the opportunity of finding significant results in the first  
5 place. Moreover, the overall credibility of a study is called into question if a sample  
6 size calculation is not reported – making the presumption possible, that the trial never  
7 underwent a proper planning phase. Without doubt a transparent sample size  
8 justification is necessary to avoid misinterpretation of study results. In summary,  
9 there is potential to improve reporting on sample size calculations in publications on  
10 AMD treatment.  
11  
12  
13  
14  
15  
16  
17

18 Literature, however, demonstrates that this tendency is by far not AMD specific. The  
19 findings of this investigation are in line with the results of other studies that have  
20 examined the quality and accuracy of the descriptions of sample size calculations in  
21 publications[4]–[9]. One study analyzed sample size calculations in publications,  
22 which had appeared in six high impact between 2005 and 2006: a total of 95% of all  
23 publications analyzed in this study provided information on the calculation of the  
24 sample size, whereas of these 43% did not report all necessary information[4].  
25 Recalculation led to a range of differences between reported and replicated sample  
26 sizes from -50% to -50%.  
27  
28  
29  
30  
31  
32  
33

34 Lee and Tse[5] examined the quality of sample size calculation in 451 RCT  
35 publications (published in December 2014 and indexed in PubMed): in 58.1% of the  
36 publications a sample size calculation was described (with recalculation having been  
37 possible for 40% of these publications). The comparison of the replicated and  
38 reported sample sizes showed a median deviation of 0 % (interquartile range -4.6%  
39 to +3%). Moreover, only 39.7% (25 out of 63) of the sample sizes were identical to  
40 the sample sizes stated in trial registers (difference: Median: 0%; interquartile range -  
41 8.1%; +15.1%). A multiple linear regression showed that journals recommending the  
42 CONSORT statement and having an impact factor published articles with more  
43 details and smaller deviations between reported and recalculated sample sizes.  
44  
45  
46  
47  
48  
49  
50

51 In other reviews 78% (66% complete)[6] and 91.7% (80.3% complete)[7] of  
52 anesthesia publications reported sample size calculations. In RCT publications from  
53 the field of dentistry and orthodontics descriptions of sample size calculations were  
54 found in only 29.3%[8] and 29.5%[9], respectively. The respective differences  
55 between the reported and replicated sample sizes were then found to range from -  
56 237.5% to 84.2%[8] and -93.3% to 60.6%[9]. Furthermore, there was also a  
57  
58  
59  
60

1  
2  
3 discrepancy between the planned and the actually recruited number of patients  
4 (recruited sample size smaller than planned sample size: 23.6%; recruited sample  
5 size larger than planned sample size: 58.4%)[11].  
6  
7

8 Some authors could demonstrate that a later year of publication had a positive effect  
9 on the completeness of sample size data[6][8][12].  
10

11 Missing sample size calculations were also found in protocols of clinical trials. From  
12 446 protocols only 42% reported all necessary elements of a sample size calculation.  
13 The replicated sample sizes were identical to the reported sample sizes in only 30 %  
14 of the trials (134)[13]. In addition it could be shown that there were also discrepancies  
15 between sample sizes in publications and protocols[14]. Another study documented  
16 that only 31 out of 71 studies (protocol/ publication) provided information on how the  
17 sample size was calculated (26 complete descriptions)[15].  
18  
19  
20  
21  
22  
23  
24

### 25 **Study limitations**

26 Evaluations and replications of sample size calculations were carried out by one  
27 consultant (ST) only (no independent parallel evaluation); however, all replications  
28 were discussed with and reviewed by a second consultant (BG) and a consensus  
29 was found by the additional review of an experienced and certified biometrician (FK)  
30 whenever deemed necessary or appropriate. A further limitation is that the  
31 assessment was not performed as a blinded procedure, i.e. the reviewers were not  
32 blinded towards the journals, publications and authors possibly having resulted in a  
33 reviewer bias. Note in addition, that several RCT author teams contributed more than  
34 one RCT reports to the 97 publications' pool, yet implying potential increase of the  
35 effect of such bias mechanisms. In addition, only a limited period of time (2004 –  
36 2014) was examined. It can be assumed that publications published after 2014 may  
37 have a higher frequency of describing sample size calculation as the journals  
38 increasingly recommend the strict use of reporting standards such as comprised by  
39 the CONSORT statement.  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50

51 From the pragmatic clinical trial investigator's perspective, this investigation's primary  
52 endpoint may furthermore have been designed overly strict for publications on RCTs  
53 on larger patient samples, as only a discrepancy of  $\pm 2$  subjects was allowed from the  
54 numerical implementation perspective. Re-analyses based on a secondary endpoint  
55 allowing for a maximum discrepancy between recalculated and reported sample  
56 sizes of  $\pm 10\%$ , however, demonstrated a similar overall tendency as observed for the  
57  
58  
59  
60

1  
2  
3 primary endpoint: 12% of 97 of the publications had a sample size calculation that  
4 reached this secondary endpoint.

5  
6 Considering the validity of reported sample size calculations, however, naturally calls  
7 for the re-assessment of the “own” sample size legitimation – yet actually having  
8 been based on the “incorrect” assumption of about 50% invalid descriptions of  
9 sample size calculation contrasted to the observed prevalence of 92%. For the  
10 confidence interval of the observed prevalence of 8% correct and complete sample  
11 size justifications, the recruited number of 97 publications must be admitted as  
12 having been chosen too small: the 95% confidence estimation of such an expected  
13 frequency would rather be based on requiring a maximum confidence interval length  
14 of, say,  $\pm 2\%$  instead of  $\pm 10\%$  (as required for the 50% prevalence assumption, see  
15 above). As a consequence, a total of 707 RCT publications would have been  
16 necessary for evaluation, yet demonstrating the essential “drawback” of sample size  
17 calculation – you only know, whether the underlying assumption and thereby the  
18 result of sample size calculations were correct after you have performed the trial.  
19 From this perspective, some of the 97 RCT publications might have omitted a sample  
20 size calculation report just for this simple reason – the initial sample size assumptions  
21 were substantially wrong. Nevertheless, transparent reporting still would encourage  
22 the reporting of the underlying assumptions and thereby explain the difference  
23 between expected and observed outcome as well as required and achieved  
24 statistical power.  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40

## 41 **Conclusion**

42 Although the CONSORT statement is available since 1996, more than half of the  
43 publications analyzed here did not report a sample size calculation. Described  
44 sample size calculations were often incorrect (calculation and practically applied  
45 sample size did not match) or incomplete (not all necessary elements were reported).  
46 This demonstrates the substantial need for improvement and, at the same time,  
47 provides constructive lines for implementation of the latter: For example, each journal  
48 could provide explicit instructions and example-illustrated guidelines for the reporting  
49 of sample size calculations. Furthermore, qualified statisticians should be involved in  
50 the planning process of a study design by means of correct sample calculations –  
51 and their active involvement in the publication process should be invoked by journals,  
52 for example, by requiring written confirmation of explicit contributions to the Methods  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 section of a submitted article. As a consequence, statisticians will be assisted in  
4 insisting, that their calculation rationale is included in any resulting publication.

5  
6 Editors and reviewers should also require each author team to provide detailed  
7 information on sample size calculations to ensure its reproducibility – at least by  
8 means of electronical supplements; the expert review of clinical articles on RCTs  
9 could, in addition, mandatorily involve qualified statistician, who could be encouraged  
10 to explicitly recalculate reported sample sizes regarding their crucial impact on the  
11 overall trial result interpretation.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

24  
25 **Funding/ Conflict of interest statement:** This systematic review was conducted by  
26 Sabrina Tulka (M. Sc. Statistics) within the framework of a full-time placement as a  
27 research assistant. This placement was funded by a 24-months research grant  
28 received from the Leonard Stinnes Foundation (internal reference: KS 11535). The  
29 study did not benefit from any industrial funding and is not subject to any financial or  
30 political conflict of interest, neither in terms of content nor with regard to the results.  
31  
32  
33  
34

35  
36 The results presented in this manuscript are part of the Doctoral Thesis of Ms. Tulka  
37 to be submitted to the Faculty of Health of Witten / Herdecke University to achieve  
38 the Doctoral degree “Dr rer medic” (Doctor of theoretical Medicine).  
39  
40

41  
42 Furthermore the results contained in this manuscript have already been presented by  
43 means of an oral presentation at the annual meeting of the German Region of the  
44 International Biometric Society (Frankfurt / Main, Germany, March 2018) and by  
45 means of a poster presentation at the annual meeting of the German Ophthalmic  
46 Surgeons (Nuremberg, Germany, June 2018), where the presentation was awarded  
47 with the 2018 poster prize of the Ophthalmic Surgeons.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57

58 **Acknowledgement:** The authors would like to express their thanks to the Leonard  
59 Stinnes Foundation, which made this research work possible in terms of the financial  
60



1  
2  
3 means for Ms. Tulka's research grant. Furthermore the authors are grateful to Drs.  
4 Christine Baulig and Stephanie Knippschild (IMBE) for providing the RCT publication  
5 pool on AMD health care, which was used for this investigation, as well as to Ms.  
6 Tara Rödter (MD) for a native speaker revision of the first draft of this manuscript.  
7  
8  
9

### 10 11 12 13 **Authorship legitimization statement / author's contribution:** 14

15  
16 S Tulka extracted the RCT publications' relevant outcome data for the sample size  
17 calculations (parallel independent evaluation), performed the sample size re-  
18 calculations, the statistical analysis of the re-calculation data and wrote this  
19 systematic review's first draft.  
20  
21  
22

23  
24 B Geis double-checked the outcome data extraction and assisted in the sample  
25 sizes' evaluation and re-calculation; furthermore she reviewed the manuscript draft.  
26  
27

28 F Krummenauer wrote the grant application to the Leonhard Stinnes Foundation,  
29 designed this investigation, extracted relevant outcome data (parallel independent  
30 evaluation) for selected RCT publications, assisted in sample size re-calculations,  
31 thoroughly revised the first draft of this manuscript and contributed major parts to the  
32 second draft version.  
33  
34  
35

36  
37 **Data sharing statement:** data may be obtained from the authors for academic  
38 purposes.  
39  
40  
41  
42  
43  
44

### 45 **References**

- 46 1 Krummenauer F, Dick B, Schwenn O, Pfeiffer N. The determination of sample  
47 size in ophthalmological trials. *British J Ophthalmol* 2002;86:946-47.
- 48 2 Moher D, Hopewell S, Schulz KF et al. CONSORT 2010 Explanation and  
49 Elaboration: updated guidelines for reporting parallel group randomised trials.  
50 *BMJ* 2010;340:c869.
- 51 3 ICH E9. Statistical principles for clinical trials. London, UK: International  
52 Conference on Harmonisation; 1998. Adopted by CPMP March 1998  
53 (CPMP/ICH/363/96).
- 54 4 Charles P, Giraudeau B, Dechartres A, et al. Reporting of sample size  
55 calculation in randomised controlled trials: review. *BMJ* 2009;338:b1732.
- 56 5 Lee PH, Tse, ACY. The quality of the reported sample size calculations in  
57 randomized controlled trials indexed in PubMed. *European Journal of Internal*  
58 *Medicine* 2017;40:16-21.  
59  
60

- 6 Chow JY, Turjstra TP, Yim E, et al. Sample size calculations for randomized clinical trials published in anesthesiology journals: a comparison of 2010 versus 2016. *Can J Anesth* 2018;65:611-618.
- 7 Abdulatif M, Mukhtar A, Obayah G. Pitfalls in reporting sample size calculation in randomized controlled trials published in leading anaesthesia journals: a systematic review. *British Journal of Anaesthesia* 2015;115:699-707.
- 8 Koletsi D, Fleming PS, Seehra J, et al. Are Sample Sizes Clear and Justified in RCTs Published in Dental Journals? *PLoS One* 2014;9:e85949.
- 9 Koletsi D, Fleming PS, Pandis N. Sample size in orthodontic randomized controlled trials: are numbers justified? *European Journal of Orthodontics* 2014;36:67-73.
- 10 Baulig C, Krummener 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.
- 11 Castellini G, Gianola S, Bonovas S, et al. Improving Power and Sample Size Calculation in Rehabilitation Trial Reports: A Methodological Assessment. *Archives of Physical Medicine and Rehabilitation* 2016;97:1195-1201.
- 12 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.
- 13 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.
- 14 Chan AW, 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:a2299.
- 15 Olberg B, Busse R, Felgentraeger K, et al. Quality of sample size estimation in trials of medical devices: high-risk devices for neurological conditions as example, *International Journal of Technology Assessment in Health Care* 2017;33:103-110.

## Legends to Figures

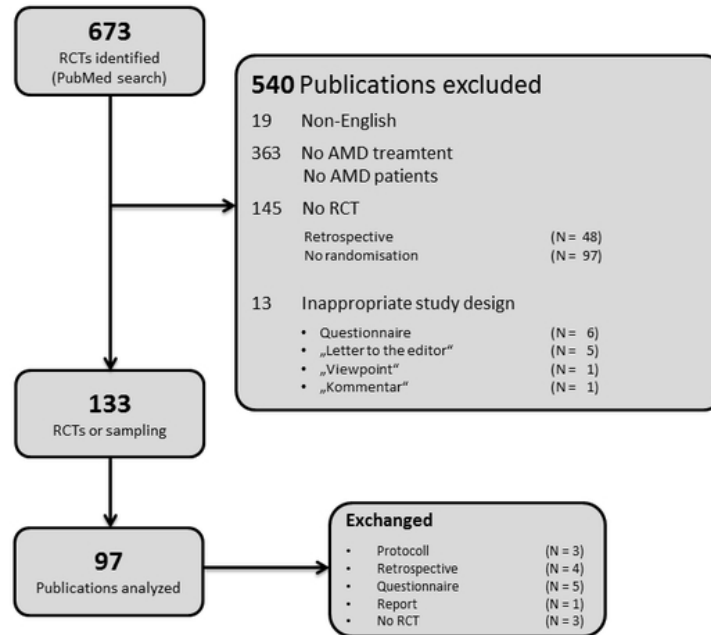
**Figure 1:** Flowchart illustrating the publication identification via PubMed (RCTs on AMD treatment published between 01/2004 and 12/2014) and the sequential process of exclusion of inappropriate publications as well as the reduction to a random sample of 97 relevant RCT publications for evaluation.

**Figure 2:** 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



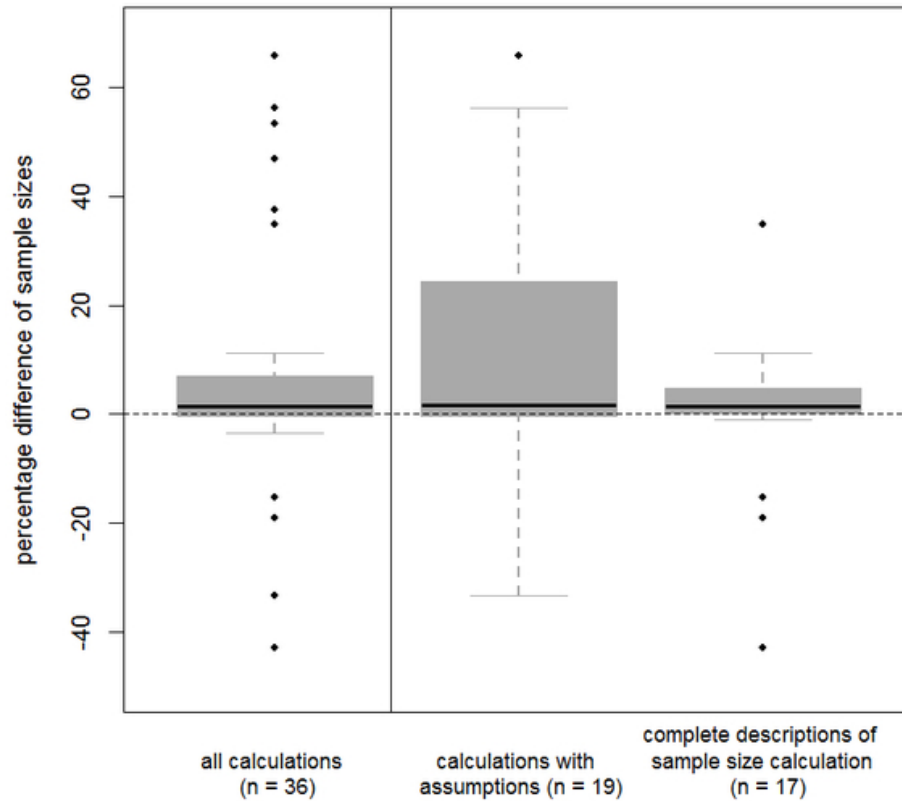
1  
2  
3 and thereby requiring assumptions or corrections for the recalculation of sample  
4 sizes (19 RCT publications); horizontal lines indicate medians and quartiles, vertical  
5 lines indicate total ranges to minimum and maximum deviations, stars indicate outlier  
6 deviation with an at least double interquartile range's deviation from the median  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only



Flowchart illustrating the publication identification via PubMed (RCTs on AMD treatment published between 01/2004 and 12/2014) and the sequential process of exclusion of inappropriate publications as well as the reduction to a random sample of 97 relevant RCT publications for evaluation.

30x22mm (600 x 600 DPI)



: 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, stars indicate outlier deviation with an at least double interquartile range's deviation from the median

26x25mm (600 x 600 DPI)

# BMJ Open

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

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2019-030312.R1
Article Type:	Original research
Date Submitted by the Author:	22-Aug-2019
Complete List of Authors:	Tulka, Sabrina; University Witten Herdecke Faculty of Health, Institute for Medical Biometry and Epidemiology, Geis, Berit; University Witten Herdecke Faculty of Health, Institute for Medical Biometry and Epidemiology Baulig, Christine; Universitat Witten/Herdecke, Medicine Knippschild, Stephanie; Universitat Witten/Herdecke, Medicine Krummenauer, Frank; University Witten Herdecke Faculty of Health, Institute for Medical Biometry and Epidemiology,
<b>Primary Subject Heading</b>:	Medical publishing and peer review
Secondary Subject Heading:	Ophthalmology
Keywords:	sample size calculation, RCT publication, transparent reporting, recalculation

SCHOLARONE™  
Manuscripts

1  
2  
3 **Validity of sample sizes in publications of randomized controlled**  
4 **trials on the treatment of age-related macular degeneration: a cross-**  
5 **sectional evaluation**  
6  
7  
8  
9  
10

11 Sabrina Tulka, Berit Geis, Christine Baulig, Stephanie Knippschild, Frank  
12 Krummenauer  
13

14  
15 University Witten / Herdecke Faculty of Health, Institute for Medical Biometry and  
16 Epidemiology, 58448 Witten, Germany  
17  
18

19  
20 Correspondence:  
21

22  
23 Sabrina Tulka, M.Sc.  
24 University Witten / Herdecke Faculty of Health,  
25 Institute for Medical Biometry and Epidemiology,  
26 Alfred Herrhausen-Straße 50, D-58448 Witten (Germany)  
27 phone: ++ 49 2302 926 471  
28 fax: ++ 49 2302 926 44785  
29 e-mail: sabrina.tulka@uni-wh.de  
30  
31  
32  
33  
34  
35

36 **Key Words:** sample size calculation, RCT publication, transparent reporting,  
37 recalculation  
38  
39  
40  
41

42 **Synopsis:** More than half of the analyzed RCT publications on the treatment of AMD  
43 did not report a transparent sample size calculation. Only 8% reported a complete and  
44 correct sample size calculation.  
45  
46  
47  
48

49 (word count: 2981)  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Abstract

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

**Methods:** A sample of 97 randomized controlled trials 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%; 16%]) publications achieved the primary endpoint. The median relative deviation between reported and recalculated sample size was 1% with a range from -43% to +66%.

**Conclusion:** Although a transparent sample size legitimization 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 legitimizations 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.

(255 words)

### Strengths and Limitations of this study:

- Validity of sample size calculations for RCT publications on the treatment of age-related macular degeneration 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 cannot be excluded.

## 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 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[1].

As a reporting guideline for publications of randomized controlled trials (RCT) the CONSORT statement[2] 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 characterizing the clinically relevant difference between the treatment samples as parameterized by the trial's primary clinical endpoint, as well as the intended levels of significance and power. In strict accordance, the 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 denotes the „International Conference on Harmonisation“ guidelines for conducting clinical trials in Europe, the United States of America and Japan). In addition, a justification for the expected effect size should be reported[3].

Despite the availability of both RCT reporting standards for longer than two decades, several investigations[4]-[9] 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 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 ophthalmologic disease, whose studies should fulfill 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 health care. The search was performed based on the following terms: “macular degeneration”, “randomized 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, SK). Publications not affiliated to AMD, without randomization, 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 electronical search strategy see Baulig et al.[10]. From this publication pool a series of 97 RCT-publications (see sample size calculation) was analyzed.

**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	11737	07:44:00
"Search (macular degeneration) AND (""2004/01/01""[Date - Publication] : ""2013/12/31""[Date - Publication])"	#10	13068	07:43:41
"Search macular degeneration"	#1	22957	07:00:29

### 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 (e.g. legitimization 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.

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 centers.

**Table 2:** Journals impact factor ranges (derived from the ISI Web of Science®) for journals having published the 97 RCT publications used for sample size evaluation, frequency of analyzed RCT publications per journal; impact factor ranges denote changes in journal impact factor over several RCT publication years.

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

### 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 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 program 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 (i.e. one of the parameters mentioned above) 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 corrected for multiplicity in the sample size calculation for a multiple trial arm comparison (e.g. by means of a Bonferroni correction); in such cases, the recalculation assumed the methodologically correct approach with regard to the study design at hand; i.e. 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 (e.g. for the  $X^2$ - 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

1  
2  
3 from previous studies mentioned in the publication at hand, or backward calculation  
4 based on the reported sample size), expected effect size (two possibilities: either the  
5 effect size from another study reported in the publication at hand was imputed, or a  
6 backward calculation was performed based on the sample size reported in the RCT-  
7 publication at hand).

### 13 **Statistical analysis and sample size calculation**

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

21 Statistical analysis of the primary endpoint was then performed by estimating its cross-  
22 sectional prevalence by means of the 95% Clopper-Pearson confidence interval.  
23 Furthermore the relative deviation [%] of the reported and recalculated sample sizes  
24 was calculated via:

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

27 To describe the distribution of these study-wise differences', medians, quartiles and  
28 ranges were estimated; non-parametric boxplots were used as a graphical  
29 presentation.

### 39 **Patient and Public Involvement**

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

## 50 **RESULTS**

### 52 **RCT publication characteristics**

54 This cross-sectional evaluation comprised 97 RCT publications from 29 journals, of  
55 which 30 (31%) were published in a journal with impact factor (IF)  $\geq 5$  at the year of  
56 publication and 67 (69%) in a journal with impact factor  $\leq 5$ . 53% of the published RCTs  
57 were multicenter trials, 51% stated industrial funding and 54% claimed the participation  
58  
59  
60

1  
2  
3 of a statistician or a statistical methods unit. In 83 of 97 (86%) RCT publications, a  
4 primary efficacy or effectiveness endpoint was examined.  
5  
6  
7

8 A total of 50 out of 97 RCT publications did not report any information on sample size  
9 calculation (95%-CI: [42%; 62%]). Eight descriptions of sample size calculation (8%;  
10 95%-CI: [4%; 16%]) were complete and reproducible, so that the underlying RCT  
11 publication achieved this investigation's primary endpoint.  
12  
13

14 The replication of reported sample size calculations was possible for 36 RCT  
15 publications (77% of the 47 publications with reported sample size legitimation; 37%  
16 of all 97 publications analyzed).  
17  
18

19 Only 17 (18% of 97) publications provided all necessary information to replicate the  
20 described sample size calculation, whereas 19 reports were incomplete or incorrect  
21 (Table 3) (however, they provided sufficient information to recalculate the sample size  
22 using values assumed from the context).  
23  
24

25 The median percentage difference between the replicated and reported sample sizes  
26 was estimated 1% (interquartile range (IQR): -1% to +5%) and the median absolute  
27 difference between the replicated and reported sample sizes was 1.50 (interquartile  
28 range (IQR): -1 to 5.25, range: -24 to + 502) for the 36 publication enabling for  
29 recalculations with or without additional assumptions due to incomplete or incorrect  
30 input data (Figure 1). Maximum deviations were -43% (reported n = 10, replicated n =  
31 7) and 66% (reported n = 261, replicated n = 763).  
32  
33

34 Among those publications reporting complete and correct input data (and thereby not  
35 requiring imputation or assumption of parameters, n = 17 publications) the median  
36 percentage difference between the reported and replicated sample sizes was again  
37 estimated 1% (interquartile range: 0% to 5%) with minimum and maximum deviations  
38 of -43% (reported n = 10, replicated n = 7) and +35% (reported n = 300, replicated n =  
39 461).  
40  
41

42 Publications in journals with an impact factor of  $\leq 5$  in the respective year of publication  
43 showed a median percentage difference of 2% (IQR: 0% - +6%), while sample size  
44 calculations in journals with an impact factor of  $> 5$  showed a median percentage  
45 difference of 0% (IQR: -1% - +3%). The median percentage difference between RCTs  
46 published before 2010 (IQR: -1% to +3%, range: -43% to 66%) and in 2010 or later  
47 (IQR: -1% to +3%, range: -33% to 56%) was 1%.  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

<b>Publications with recalculation via imputed values</b>	
<b>Wrong values</b>	
<b>Power</b>	5
<b>Level of significance</b>	3
<b>Effect size</b>	2
<b>Method (statistical test)</b>	1
<b>Missing values</b>	
<b>Power</b>	-
<b>Level of significance</b>	2
<b>Effect size</b>	12
<b>Publications without recalculation</b>	
<b>Pilot study (1 with budgetary limitations)</b>	2
<b>No sample size reported</b>	1
<b>Incorrect or inconclusive planning (poweranalysis)</b>	8

## DISCUSSION

This cross-sectional investigation demonstrated a notable lack of methodological transparency and correctness of sample size calculations in AMD RCT-publications (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  $\pm 2$  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 analyzed 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 hypothesized above: the transparent reporting of a sample size

1  
2  
3 calculation is an important tool for assessing whether a study was planned carefully  
4 and had the opportunity of finding significant results in the first place. Moreover, the  
5 overall credibility of a study is called into question if a sample size calculation is not  
6 reported – making the presumption possible, that the trial never underwent a proper  
7 planning phase. Without doubt a transparent sample size justification is necessary to  
8 avoid misinterpretation of study results. In summary, there is potential to improve  
9 reporting on sample size calculations in publications on AMD treatment. A logistic  
10 regression did not reveal factors (IF, funding, year of publication) clearly associated  
11 with a study's chance of reaching our study's primary endpoint).  
12  
13  
14  
15  
16  
17  
18  
19

20 Literature, however, demonstrates that this tendency is by far not AMD specific. The  
21 findings of this investigation are in line with the results of other studies that have  
22 examined the quality and accuracy of the descriptions of sample size calculations in  
23 publications [4]–[9]. One study analyzed sample size calculations in publications,  
24 which had appeared in six high impact between 2005 and 2006: a total of 95% of all  
25 publications analyzed in this study provided information on the calculation of the  
26 sample size, whereas of these 43% did not report all necessary information[4].  
27 Recalculation led to a range of differences between reported and replicated sample  
28 sizes from -50% to -50%.  
29  
30  
31  
32  
33  
34  
35

36 Lee and Tse [5] examined the quality of sample size calculation in 451 RCT  
37 publications (published in December 2014 and indexed in PubMed): in 58.1% of the  
38 publications a sample size calculation was described (with recalculation having been  
39 possible for 40% of these publications). The comparison of the replicated and reported  
40 sample sizes showed a median deviation of 0 % (interquartile range -4.6% to +3%).  
41 Moreover, only 39.7% (25 out of 63) of the sample sizes were identical to the sample  
42 sizes stated in trial registers (difference: Median: 0%; interquartile range -8.1%;  
43 +15.1%). A multiple linear regression showed that journals recommending the  
44 CONSORT statement and having an impact factor published articles with more details  
45 and smaller deviations between reported and recalculated sample sizes.  
46  
47  
48  
49  
50  
51  
52

53 In other reviews 78% (66% complete) [6] and 91.7% (80.3% complete)[7] of anesthesia  
54 publications reported sample size calculations. In RCT publications from the field of  
55 dentistry and orthodontics descriptions of sample size calculations were found in only  
56 29.3%[8] and 29.5%[9], respectively. The respective differences between the reported  
57 and replicated sample sizes were then found to range from -237.5% to 84.2%[8] and -  
58  
59  
60



1  
2  
3 93.3% to 60.6%[9]. Furthermore, there was also a discrepancy between the planned  
4 and the actually recruited number of patients (recruited sample size smaller than  
5 planned sample size: 23.6%; recruited sample size larger than planned sample size:  
6 58.4%)[11].  
7  
8

9  
10 Some authors could demonstrate that a later year of publication had a positive effect  
11 on the completeness of sample size data [6][8][12].  
12

13 Missing sample size calculations were also found in protocols of clinical trials. From  
14 446 protocols only 42% reported all necessary elements of a sample size calculation.  
15 The replicated sample sizes were identical to the reported sample sizes in only 30 %  
16 of the trials (134)[13]. In addition it could be shown that there were also discrepancies  
17 between sample sizes in publications and protocols[14]. Another study documented  
18 that only 31 out of 71 studies (protocol/ publication) provided information on how the  
19 sample size was calculated (26 complete descriptions)[15].  
20  
21  
22  
23  
24  
25  
26

### 27 **Study limitations**

28  
29 Evaluations and replications of sample size calculations were carried out by one  
30 consultant (ST) only (no independent parallel evaluation); however, all replications  
31 were discussed with and reviewed by a second consultant (BG) and a consensus was  
32 found by the additional review of an experienced and certified biometrician (FK)  
33 whenever deemed necessary or appropriate. A further limitation is that the assessment  
34 was not performed as a blinded procedure, i.e. the reviewers were not blinded towards  
35 the journals, publications and authors possibly having resulted in a reviewer bias (for  
36 example in rating a value as wrong). Note in addition, that several RCT author teams  
37 contributed more than one RCT reports to the 97 publications' pool, yet implying  
38 potential increase of the effect of such bias mechanisms. In addition, only a limited  
39 period of time (2004 – 2014) was examined. It can be assumed that publications  
40 published after 2014 may have a higher frequency of describing sample size  
41 calculation as the journals increasingly recommend the strict use of reporting  
42 standards such as comprised by the CONSORT statement. This period of time was  
43 chosen as this project was an add-on to a project on search strategies [16]. A follow-  
44 up project on publications after 2014 is planned.  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55

56 From the pragmatic clinical trial investigator's perspective, this investigation's primary  
57 endpoint may furthermore have been designed overly strict for publications on RCTs  
58 on larger patient samples, as only a discrepancy of  $\pm 2$  subjects was allowed from the  
59  
60

1  
2  
3 numerical implementation perspective. Re-analyses based on a secondary endpoint  
4 allowing for a maximum discrepancy between recalculated and reported sample sizes  
5 of  $\pm 10\%$ , however, demonstrated a similar overall tendency as observed for the  
6 primary endpoint: 12% of 97 of the publications had a sample size calculation that  
7 reached this secondary endpoint.  
8  
9

10  
11 Considering the validity of reported sample size calculations, however, naturally calls  
12 for the re-assessment of the “own” sample size legitimization – yet actually having been  
13 based on the “incorrect” assumption of about 50% invalid descriptions of sample size  
14 calculation contrasted to the observed prevalence of 92%. For the confidence interval  
15 of the observed prevalence of 8% correct and complete sample size justifications, the  
16 recruited number of 97 publications must be admitted as having been chosen too small:  
17 the 95% confidence estimation of such an expected frequency would rather be based  
18 on requiring a maximum confidence interval length of, say,  $\pm 2\%$  instead of  $\pm 10\%$  (as  
19 required for the 50% prevalence assumption, see above). As a consequence, a total  
20 of 707 RCT publications would have been necessary for evaluation, yet demonstrating  
21 the essential “drawback” of sample size calculation – you only know, whether the  
22 underlying assumption and thereby the result of sample size calculations were correct  
23 after you have performed the trial. From this perspective, some of the 97 RCT  
24 publications might have omitted a sample size calculation report just for this simple  
25 reason – the initial sample size assumptions were substantially wrong. Nevertheless,  
26 transparent reporting still would encourage the reporting of the underlying assumptions  
27 and thereby explain the difference between expected and observed outcome as well  
28 as required and achieved statistical power.  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44

## 45 **Conclusion**

46 Although the CONSORT statement is available since 1996, more than half of the  
47 publications analyzed here did not report a sample size calculation. Described sample  
48 size calculations were often incorrect (calculation and practically applied sample size  
49 did not match) or incomplete (not all necessary elements were reported). This  
50 demonstrates the substantial need for improvement and, at the same time, provides  
51 constructive lines for implementation of the latter: For example, each journal could  
52 provide explicit instructions and example-illustrated guidelines for the reporting of  
53 sample size calculations. Furthermore, qualified statisticians should be involved in the  
54 planning process of a study design by means of correct sample calculations – and their  
55  
56  
57  
58  
59  
60



1  
2  
3 active involvement in the publication process should be invoked by journals, for  
4 example, by requiring written confirmation of explicit contributions to the Methods  
5 section of a submitted article. As a consequence, statisticians will be assisted in  
6 insisting, that their calculation rationale is included in any resulting publication.  
7  
8

9  
10 Editors and reviewers should also require each author team to provide detailed  
11 information on sample size calculations to ensure its reproducibility – at least by means  
12 of electronical supplements; the expert review of clinical articles on RCTs could, in  
13 addition, mandatorily involve qualified statistician, who could be encouraged to  
14 explicitly recalculate reported sample sizes regarding their crucial impact on the overall  
15 trial result interpretation.  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27

28 **Funding statement:** This work was supported by Leonard Stinnes Foundation grant  
29 number (internal reference: KS 11535).  
30  
31  
32  
33  
34

35 **Conflict of interest statement:** This systematic review was conducted by Sabrina  
36 Tulka (M. Sc. Statistics) within the framework of a full-time placement as a research  
37 assistant. This placement was funded by a 24-months research grant received from  
38 the Leonard Stinnes Foundation (internal reference: KS 11535). The study has no  
39 political conflict of interest, neither in terms of content nor with regard to the results.  
40  
41  
42  
43

44 The results presented in this manuscript are part of the Doctoral Thesis of Ms. Tulka  
45 to be submitted to the Faculty of Health of Witten / Herdecke University to achieve the  
46 Doctoral degree “Dr rer medic” (Doctor of theoretical Medicine).  
47  
48  
49

50 Furthermore the results contained in this manuscript have already been presented by  
51 means of an oral presentation at the annual meeting of the German Region of the  
52 International Biometric Society (Frankfurt / Main, Germany, March 2018) and by means  
53 of a poster presentation at the annual meeting of the German Ophthalmic Surgeons  
54 (Nuremberg, Germany, June 2018), where the presentation was awarded with the  
55 2018 poster prize of the Ophthalmic Surgeons.  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8 **Acknowledgement:** The authors would like to express their thanks to the Leonard  
9 Stinnes Foundation, which made this research work possible in terms of the financial  
10 means for Ms. Tulka's research grant. Furthermore to Ms. Tara Rödter (MD) for a  
11 native speaker revision of the first draft of this manuscript.  
12  
13  
14  
15  
16  
17

### 18 **Authorship legitimisation statement / author's contribution:**

19  
20  
21 S Tulka extracted the RCT publications' relevant outcome data for the sample size  
22 calculations (parallel independent evaluation), performed the sample size re-  
23 calculations, the statistical analysis of the re-calculation data and wrote this systematic  
24 review's first draft.  
25  
26  
27

28  
29 B Geis double-checked the outcome data extraction and assisted in the sample sizes'  
30 evaluation and re-calculation; furthermore she reviewed the manuscript draft.  
31  
32

33 C Baulig carried out the publication search, excluded inappropriate articles and revised  
34 the manuscript.  
35  
36

37 S Knippschild carried out the publication search, excluded inappropriate articles and  
38 revised the manuscript.  
39  
40

41 F Krummenauer wrote the grant application to the Leonhard Stinnes Foundation,  
42 designed this investigation, extracted relevant outcome data (parallel independent  
43 evaluation) for selected RCT publications, assisted in sample size re-calculations,  
44 thoroughly revised the first draft of this manuscript and contributed major parts to the  
45 second draft version.  
46  
47  
48  
49

50  
51 **Data sharing statement:** data may be obtained from the authors for academic  
52 purposes.  
53  
54  
55  
56

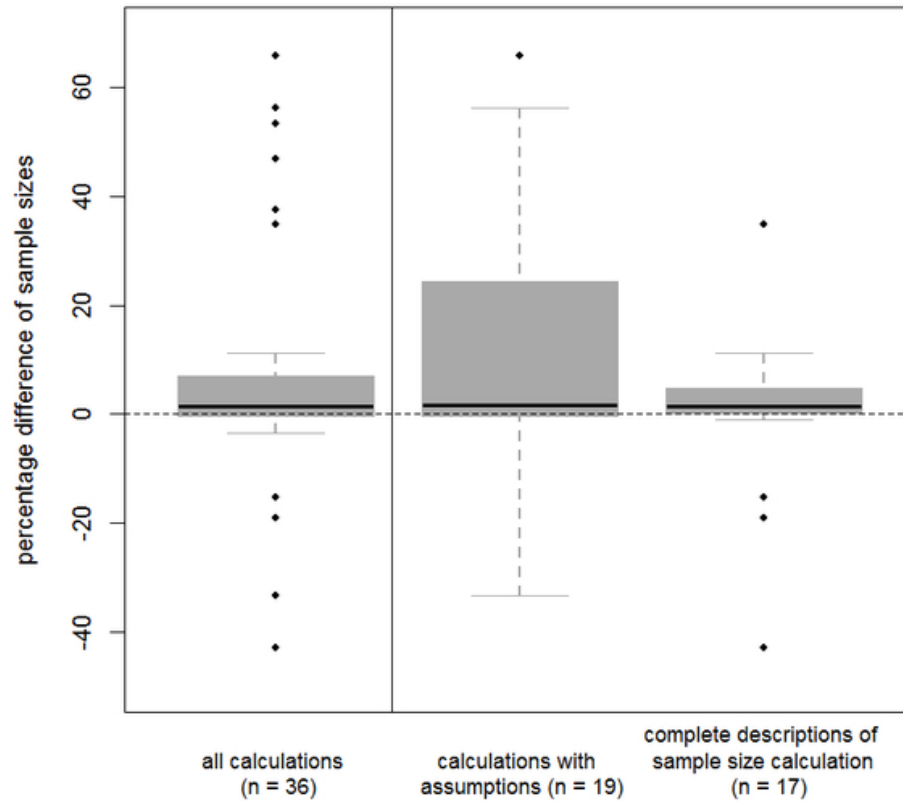
### 57 **References**

- 58  
59  
60 1 Krummenauer F, Dick B, Schwenn O, Pfeiffer N. The determination of sample size in ophthalmological trials. *British J Ophthalmol* 2002;86:946-47.

- 1
- 2
- 3 2 Moher D, Hopewell S, Schulz KF et al. CONSORT 2010 Explanation and
- 4 Elaboration: updated guidelines for reporting parallel group randomised trials.
- 5 BMJ 2010;340:c869.
- 6
- 7 3 ICH E9. Statistical principles for clinical trials. London, UK: International
- 8 Conference on Harmonisation; 1998. Adopted by CPMP March 1998
- 9 (CPMP/ICH/363/96).
- 10
- 11 4 Charles P, Giraudeau B, Dechartres A, et al. Reporting of sample size
- 12 calculation in randomised controlled trials: review. BMJ 2009;338:b1732.
- 13
- 14 5 Lee PH, Tse, ACY. The quality of the reported sample size calculations in
- 15 randomized controlled trials indexed in PubMed. European Journal of Internal
- 16 Medicine 2017;40:16-21.
- 17
- 18 6 Chow JY, Turjstra TP, Yim E, et al. Sample size calculations for randomized
- 19 clinical trials published in anesthesiology journals: a comparison of 2010
- 20 versus 2016. Can J Anesth 2018;65:611-618.
- 21
- 22 7 Abdulatif M, Mukhtar A, Obayah G. Pitfalls in reporting sample size calculation
- 23 in randomized controlled trials published in leading anaesthesia journals: a
- 24 systematic review. British Journal of Anaesthesia 2015;115:699-707.
- 25
- 26 8 Koletsi D, Fleming PS, Seehra J, et al. Are Sample Sizes Clear and Justified
- 27 in RCTs Published in Dental Journals? PLoS One 2014;9:e85949.
- 28
- 29 9 Koletsi D, Fleming PS, Pandis N. Sample size in orthodontic randomized
- 30 controlled trials: are numbers justified? European Journal of Orthodontics
- 31 2014;36:67-73.
- 32
- 33 10 Baulig C, Krummenauer F, Geis B, et al. Reporting quality of randomised
- 34 controlled trials abstracts on age-related macular degeneration health care: a
- 35 cross-sectional quantification of the adherence to CONSORT abstract
- 36 reporting recommendations. BMJ Open 2018;8:e021912.
- 37
- 38 11 Castellini G, Gianola S, Bonovas S, et al. Improving Power and Sample Size
- 39 Calculation in Rehabilitation Trial Reports: A Methodological Assessment.
- 40 Archives of Physical Medicine and Rehabilitation 2016;97:1195-1201.
- 41
- 42 12 Martin J, Taljaard M, Girling A, et al. Systematic review finds major
- 43 deficiencies in sample size methodology and reporting for stepped-wedge
- 44 cluster randomised trials. BMJ Open 2016;6:e010166.
- 45
- 46 13 Clark T, Berger U, Mansmann U. Sample size determinations in original
- 47 research protocols for randomised clinical trials submitted to UK research
- 48 ethics committees: review. BMJ 2013;346:f1135.
- 49
- 50 14 Chan AW, Hróbjartsson A, Jørgensen KJ, et al. Discrepancies in sample size
- 51 calculations and data analyses reported in randomised trials: comparison of
- 52 publications with protocols. BMJ 2008;337:a2299.
- 53
- 54 15 Olberg B, Busse R, Felgentraeger K, et al. Quality of sample size estimation in
- 55 trials of medical devices: high-risk devices for neurological conditions as
- 56 example, International Journal of Technology Assessment in Health Care
- 57 2017;33:103-110.
- 58
- 59 16 Baulig C, Niemeyer L, Krummenauer F, Knippschild S (2019) Comparison of
- 60 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.

## Legends to Figures

**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, stars indicate outlier deviation with an at least double interquartile range's deviation from the median.



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, stars indicate outlier deviation with an at least double interquartile range's deviation from the median

26x25mm (600 x 600 DPI)