Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols

Rahul Mhaskar[1], **Benjamin Djulbegovic**[1,2], **Anja Magazin**[2], **Heloisa P. Soares**[2], and **Ambuj Kumar**[1]

[1]Division and Center for Evidence Based Medicine, Department of Internal Medicine, Clinical and Translational Science Institute, University of South Florida

[2]Moffitt Cancer Center, Tampa, FL

Abstract

Objectives—To assess whether reported methodological quality of randomized controlled trials (RCTs) reflect the actual methodological quality, and to evaluate the association of effect size (ES) and sample size with methodological quality.

Study design—Systematic review

Setting—Retrospective analysis of all consecutive phase III RCTs published by 8 National Cancer Institute Cooperative Groups until year 2006. Data were extracted from protocols (actual quality) and publications (reported quality) for each study.

Results—429 RCTs met the inclusion criteria. Overall reporting of methodological quality was poor and did not reflect the actual high methodological quality of RCTs. The results showed no association between sample size and actual methodological quality of a trial. Poor reporting of allocation concealment and blinding exaggerated the ES by 6% (ratio of hazard ratio [RHR]: 0.94, 95%CI: 0.88, 0.99) and 24% (RHR: 1.24, 95%CI: 1.05, 1.43), respectively. However, actual quality assessment showed no association between ES and methodological quality.

Conclusion—The largest study to-date shows poor quality of reporting does not reflect the actual high methodological quality. Assessment of the impact of quality on the ES based on reported quality can produce misleading results.

Introduction

Randomized controlled trials (RCTs) are considered the most reliable method to assess the efficacy of competing interventions. Well designed and conducted RCTs and meta-analyses of RCTs are essential to ascertain whether new treatments offer small or moderate, but worthwhile, benefits.[1, 2] Users of research evidence (i.e., physicians, patients and policy-

Corresponding Author: Ambuj Kumar, MD, MPH, Assistant Professor and Deputy Director, Division and Center for Evidence Based Medicine and Outcomes Research, Department of Internal Medicine, College of Medicine, University of South Florida, Clinical and Translational Science Institute, 12901 Bruce B Downs Blvd, MDC 27, Tampa, FL 33612, USA, Phone: 813-396-9194, Fax: 813-905-8909, akumar1@hsc.usf.edu.

makers) make decisions on the basis of their confidence in the methodological quality of data presented in the publications. A large body of empirical evidence shows that biased results from poorly designed and reported RCTs can mislead decision making in health care. [3] Accordingly, assessment of methodological quality of RCTs is crucial for informed healthcare decision making which has also been emphasized in a recent report by Institute of Medicine[4] However, whether, published reports of RCTs truly reflect the actual methodological quality of RCTs has been assessed in only 3 cohorts of RCTs to date. [5–7] The first study by Soares et. al. assessed the publications and protocols of 59 phase III RCTs conducted by Radiation Therapy Oncology Group in the US.[5] The study concluded that published reports do not reflect the actual superior methodological quality of RCTs.[5] Similarly, a study by Hill et. al. assessed the published methodological quality of 40 RCTs in rheumatology and reported that published results do not represent the true methodological quality.[6] Similarly, the study by Devereaux and colleagues assessed publications of randomly selected 105 RCTs and concluded that authors failed to report the allocation concealment and blinding procedures conducted in these RCTs.[7] However, if these findings are applicable to other cohort of RCTs is not known. Moreover, what matters to the user of randomized evidence is not methodological quality *per se*, but whether the quality affects treatment effect size. To date, the relative impact of actual methodological quality of conduct versus methodological quality of reporting on treatment effect size in RCTs has not been studied. Some investigators have also postulated that RCTs with larger sample size may be associated with better methodological quality and that the larger sample size might have an impact on outcomes (i.e. in terms of results favoring standard or experimental treatment).[8, 9] Nevertheless, the evidence on the association of sample size of a RCT and methodological quality and associated outcomes is conflicting.[8–10] While some studies found an association between sample size and methodological quality and resulting outcomes [8, 9], others found no such association.[10] In addition, the association of sample size with methodological quality of RCTs and resulting outcomes has not been evaluated in oncology RCTs, which is the subject of this paper.

The objectives of the current study are to assess 1) whether published methodological quality of a RCT truly reflect the actual methodological quality as reflected in the study protocol; 2) assess the impact of methodological quality of reporting and actual methodological quality on treatment effect size; 3) association of RCT sample size with methodological quality and outcomes.

Methods

The objectives of this study were addressed using the systematic review methodology. The study plan was specified *a priori* in a protocol.[11]

Inclusion criteria

All consecutive terminated phase III RCTs conducted by 8 National Cancer Institute (NCI) sponsored Cooperative Groups (COG) published until year 2006 and for which both full protocol and publications were available are included in the study. The 8 COGs are Children's Oncology Group, National Surgical Adjuvant Breast and Bowel Project, Radiation Therapy Oncology Group, North Central Cancer Treatment Group, Gynecology Oncology Group, Eastern Cooperative Group, Cancer and Leukemia Group B and Southwest Cooperative Group.

Data sources

A list of all consecutive RCTs and associated study protocols were obtained from respective COGs which also provided a matching publication associated with each RCT.

Study selection

An initial list of all consecutive RCT and associated protocols was reviewed independently by 2 authors for eligibility. Studies not meeting the inclusion criteria were excluded.

Data extraction

Two reviewers independently extracted data using a standardized data extraction form. Data was extracted from both, protocol and matching publication for methodological quality domains relevant to minimizing bias and random error for each included RCT according to the methods recommended by Cochrane Collaboration.[12, 13] The extracted data from protocol and publication were then classified according to their source as: a) data reported only in publication b) data reported only in protocol C) data reported in either protocol or publication d) data reported neither in protocol nor in publication. Our "final assessment" of reporting of methodological quality was based on data from *either* protocol or the publication. We also extracted data on outcome of overall survival (hazard ratio (HR) and associated 95% confidence intervals) from each RCT which was used as the estimate of treatment effect size (ES). When direct extraction was not feasible, methods by Tierney et al[14] were used.

Statistical analysis

We used descriptive statistics for reporting of methodological quality domains in protocols and publications. Association between methodological quality (reported versus actual) with ES was conducted using standard meta-epidemiologic methods. [15] Briefly, we computed the combined treatment ES estimates separately in trials with and without the methodological quality domain interest (e.g. inadequate or unclear allocation concealment) to calculate the ratio of hazard ratios (RHR) and 95% confidence interval (CI). The effect of methodological quality domains in RCTs on sample size was tested using Kruskal–Wallis one-way analysis of variance test.[16] To test the association of RCT sample size and outcomes, we analyzed the correlation between the RCT sample size and treatment effect size (HR) using Spearman rank-correlation test.[17] The Kruskal–Wallis one-way analysis of variance test [16] was used to test the equality of median sample size across three possible outcomes: RCTs favoring new treatment (defined as upper limit of overall survival HR 95% CI less than one), RCTs favoring standard treatment (defined as lower limit of overall survival HR 95% CI greater than 1) and RCTs favoring none of the treatments (defined as overall survival HR 95% CI including 1).

Results

Between years 1968 and 2006, the NCI cooperative groups conducted 622 RCTs involving 781 comparisons. Out of 622 studies, protocols or publications were not available for 276 comparisons (117 RCTs) and therefore were not included in the final analysis. Of the remaining 505 comparisons, 76 (15%, 76/505) shared the protocols and were therefore excluded, resulting in 429 unique RCTs (enrolling 158,000 patients) which were used in the final analysis (Figure 1). The publication rate for the 429 RCTs was 98% (421/429).

Reporting of factors associated with risk for bias (protocol versus publications)

The evaluation of key methodological domains associated with risk for bias is illustrated in Figure 2. While, 39% (165/429) of RCTs employed an adequate method for randomization sequence generation as stated in the protocol, only 23% (98/429) reported doing so in the publications. Treatment allocation was adequately concealed and specified in protocols in 95% (409/429) of RCTs but only 24% (101/429) reported correctly in the publications. The procedure for blinding was applicable to 35 RCTs only. Of these 35 RCTs the procedure for

blinding was adequately reported in 91% (32/35) of protocols versus 43% (15/35) of publications. Thirty three percent (141/429) of RCTs reported expected drop-out rate in the protocols and 97% (418/426) of RCTs stated the drop-out rate in publications as well.

The plan to follow the intention to treat principle was reported in protocols in 21% (91/428) versus 91% (392/428) of RCTs actually reporting it in publications. However, out of these 392 RCTs reporting ITT analysis, only 15 % (63/392) of RCTs mentioned the exact "intention to treat" phrase in the publications. In the remaining 77% (329/392) of RCTs it was clear (by matching randomized and analyzed populations) that ITT principle was used but these RCTs did not state the exact phrase or provided an explanation that analysis were based on ITT principle. Eighty two percent (351/429) of RCTs had active treatment as comparator, 7% (30/429) of RCTs employed placebo as comparator, 10% (45/429) of RCTs had no treatment as comparator and 1% (3/429) of RCTs had active treatment and add-on placebo as comparator.

Reporting of factors associated with risk for random error (protocol versus publications)

The expected difference in the primary outcome was stated *a priori* in the protocol of 96% (414/429) of RCTs, but was mentioned in the publication for 43% (186/429) of RCTs only. The α and β errors were pre-specified in the protocol for 93% (398/426) and 95% (406/426) of RCTs, respectively. However, only 32% (138/426) of RCTs reported α error and 41% (174/426) of RCTs reported β errors in the publications. *A priori* sample size calculations were performed in 98% (421/429) of RCTs while only 40% (172/429) of RCTs reported of having done so (figure 3).

Association of methodological quality of RCTs and treatment effect size

There was no statistically significant difference between associations of ES and methodological quality of reporting for the following domains: adequacy of randomization sequence generation, description of drop outs, intention to treat analysis, pre specification of alpha and beta errors. However, on average, poorly reported allocation concealment exaggerated the ES by 6% (RHR: 0.94, 95% CI: 0.88–0.99). Also, poorly reported blinding inflated the ES by 24% (RHR: 1.24, 95% CI: 1.05–1.43; figure 4).

Nonetheless, when the data from either protocol or the publication was taken into consideration in our final assessment of methodological quality, there was no statistically significant association between ES and any of methodological quality domains (figure 4).

Association of sample size and methodological quality of RCTs

The distribution of median sample size of RCTs was similar across RCTs with adequate versus inadequate description of generation of randomization sequence as reported in publications only (p value=0.28). Similarly, there was no difference in median sample size across trials reporting adequate versus inadequate allocation concealment (p value=0.09), and adequate versus inadequate description of drop outs (p value=0.56) in the publications. Furthermore, there was no difference in median sample size in trials with pre specified α error (p value=0.14), and β error (p value=0.23) compared with RCTs which did not specify α error and β error in publications. However, RCTs which adequately described the blinding procedures enrolled more patients (median: 234; range: 47–1387) compared with RCTs which did not described the blinding procedures (median: 123; range: 45–18882) (p value = 0.05). Choice of the comparator was also associated with the median RCT sample size (p value < 0.001). *(see* table 1) However, taking into account the final assessment of methodological quality of RCT, regardless of the source (protocol or publication), the results showed that variation in sample size of a RCT was not associated with methodological quality *(see* table 2).

Association of sample size and effect size estimates of RCTs

There was a statistically significant negative correlation between RCT sample size and absolute treatment effect size for the outcome of overall survival (Rho = −0.147, p value = 0.006). However, the distribution of RCT sample size was similar across RCTs favoring new treatment, RCTs favoring standard treatment and RCTs favoring none of the treatments (p value = 0.13).

Discussion

The practice of medicine is informed by new research findings, and physicians incorporate the evidence into practice after assessing the quality of evidence as reported in the peer-reviewed publications. That is, quality of reporting is vital because the users of research evidence (i.e., physicians, patients, policy- makers) make decisions on the basis of their confidence in the accuracy of a given research paper. Our findings show that the quality of reporting of the NCI cooperative groups is rather poor. Publications often omit important methodological features which are critical to decision making. However, the findings also suggest that poor reporting of RCTs does not correlate with actual superior methodological quality of trials. Our study was limited to the cohort of RCTs for which both the protocol and publication were available. As a result, we excluded 117 RCTs (276 comparisons) from our analysis (figure 1). However, the methodological quality of reporting in these 117 RCTs was similar to the RCTs included in our analysis (data not shown). Also, details related to generation of randomization sequence were reported poorly in protocol and publications which may be an artifact of the strict definitions we applied towards the assessment of methodological quality. Additionally, NCI cooperative groups with a centralized mechanism with common and standardized methods for randomization might find it unimportant to report these details in protocols as well.

Our study results also show that the association of methodological quality of reporting and treatment effect is spurious as evident from the actual high methodological quality of conduct of these RCTs. That is, we found a statistically significant association between treatment effect size and reporting (published data only) of methodological quality domains of allocation concealment and blinding. That is, poor reporting of allocation concealment and blinding inflates the ES. However, assessment of methodological quality using data from protocols and publications showed no impact of methodological quality on treatment effect size. Results also showed that for the cohort of RCTs conducted by the NCI cooperative groups the published methodological quality was also associated with trial sample size. In contrast, taking into account the actual methodological quality the association between median sample size and quality domains was non-existent. That is, we found a statistically significant variation in the distribution of median RCT sample size based on the reporting (published data only) of methodological quality domains such as choice of the comparator (placebo) and description of blinding procedure (table 1). However, assessment of methodological quality using data from protocol and publications showed no variation in the sample size across RCTs (table 2). The results further emphasize the importance of including protocols and publications for assessment of methodological quality.

Our findings are in line with previous research on the topic. [5] [2, 6, 18] However, except one study that focused on radiation oncology trials [5] none of the previous study used study protocols for assessment of actual methodological quality i.e. reported a comparison of the quality of reporting with the methods specified in the original research protocols. The current study is also largest to date on the subject.

It has been previously shown that low methodological quality of RCTs inflates the treatment effect. For example, Colditz et. al. found that double-blind RCTs had smaller ES than nonblinded trials.[19] Similarly, Schulz et. al. reported that inadequate allocation concealment accounted for a substantial increase in ES.[20] Allocation concealment has shown the most consistent associations with treatment effect sizes.[21] However, all of these studies have assessed the methodological quality of reporting only. The relative impact of the actual methodological quality of conduct versus the methodological quality of reporting on ES has not been studied. Our results show that the assessment of the impact of methodological quality on the ES based only on the quality of reporting can produce spurious results. These findings are important for meta-epidemiologic research and further development of quality assessment tools. Researchers may need to revise the current strategy of assessment of methodological quality based on reporting only. In instances where the study protocol is available, the assessment of methodological quality of conduct may augment the overall assessment of methodological quality of these studies.

Our results also showed that median sample size of a RCT was also associated with reported methodological quality domains of blinding and the choice of the comparator (placebo versus active treatment). However, there was no association when actual methodological quality of trials as reported in the protocols was assessed. We are aware of one study by Singh et. al. which concluded that RCT sample size was related to methodological quality domains of blinding and use of ITT principle of data analysis.[9] This study also reported that sample size was an independent predicator of positive result in a RCT. In contrast, our findings suggest that the reported RCT sample size is not associated with outcomes (P = 0.11). The discrepancy in findings between the study by Singh et al. and ours could be attributed to the facts that this study reviewed only arhtroplasty RCTs and dichotomized RCT sample size as large (100 patients) versus small sample size (100 patients).[9] Additionally, the study by Singh et. al. based their analysis on published methodological quality only and did not have access to study protocols to determine the actual methodological quality of RCTs.[9]

In summary, we report the first comprehensive, formal investigation of the methodological quality of RCTs conducted by the NCI cooperative groups. While the results show that, RCTs conducted by the NCI cooperative groups are of high quality the published reports have deficiencies in their description of the actual methods used in the RCTs. It is important to note that these RCTs were of high quality even before publication of the Consolidated Standards of Reporting of Trials (CONSORT) statement in 1996.[22] Our findings indicate that although investigators in NCI cooperative groups were attentive to the critical aspects of the design and conduct of RCTs, they were less aware of the need to report these features. Given that physicians, policy-makers, guidelines panels, systematic reviewers and other users of evidence rely on published reports only, this is important findings that can be immediately rectified by the NCI "officially" adopting the CONSORT reporting guidelines. Our study also highlights the pitfall of current practices of appraisal of based on methodological quality of reporting only without considering the actual methodological quality of conduct evident from research protocols.

These findings underline the overall need for enhanced adherence to the revised CONSORT statement by both authors and journal editors which will allow transparent evaluation of RCTs.[23] The results also underscore the importance of publication of RCT protocols in the public domain. Publication of study protocols will not only help decision makers in interpreting the results from RCTs transparently and efficiently, but might also enhance prospects for collaboration thereby decreasing replication of research effort, improve RCT recruitment, and reduce bias in the reporting.[24–27] These findings are also important for further development of the methodological quality appraisal tools such as Cochrane "risk of

bias" algorithm and their application in assessment of the actual methodological quality versus quality of reporting.

Acknowledgments

References

1. Moher D, Ba P, Jones D, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analysis? Lancet. 1998; 352:609–13. [PubMed: 9746022]

2. Huwiler-Muntener KMD, Juni PMD, Junker CMDM, Egger MMDM. Quality of Reporting of Randomized Trials as a Measure of Methodologic Quality. JAMA PEER REVIEW CONGRESS IV. 2002; 287(21):2801–4. http://www.jama.com.

3. Moher D, Jones A, Lepage L. CONSORT Group (Consolidated Standards of Reporting Trials). Use of the CONSORT statement and quality of reports of randomized trials. A comparative before-and-after evaluation. JAMA. 2001; 285:1992–5. [PubMed: 11308436]

4. IOM. Finding what works in health care: statndards for systematic reviews. Washington DC: The national academics press; 2011.

5. Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, Djulbegovic B. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. Bmj. 2004 Jan 3; 328(7430):22–4. [PubMed: 14703540]

6. Hill CL, LaValley MP, Felson DT. Discrepancy between published report and actual conduct of randomized clinical trials. J Clin Epidemiol. 2002 Aug; 55(8):783–6. [PubMed: 12384192]

7. Devereaux PJ, Choi PT, El-Dika S, Bhandari M, Montori VM, Schunemann HJ, Garg AX, Busse JW, Heels-Ansdell D, Ghali WA, Manns BJ, Guyatt GH. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. J Clin Epidemiol. 2004 Dec; 57(12):1232–6. [PubMed: 15617948]

8. Kjaergard LL, Nikolova D, Gluud C. Randomized clinical trials in HEPATOLOGY: predictors of quality. Hepatology. 1999 Nov; 30(5):1134–8. [PubMed: 10534332]

9. Singh JA, Murphy S, Bhandari M. Trial sample size, but not trial quality, is associated with positive study outcome. Journal of Clinical Epidemiology. 63(2):154–62. [PubMed: 19716266]

10. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med. 2001 Dec 4; 135(11):982–9. [PubMed: 11730399]

11. Djulbegovic, B. NIH/ORI 2005–2007. 5 R01 NS052956-02: Evaluation of quality of clinical trials' records.

12. The Cochrane Collaboration. Cochrane Hand Search Manual. 1998. http://hiru.mcmaster.ca.cochrane/registry/handsearch/hsmpt1.htm

13. Egger, MSD.; Altman, D. Systematic reviews in health care. Meta-analysis in context. 2. London: BMJ; 2001.

14. Tierney J, Stewart L, Ghersi D, Burdett S, Sydes M. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials. 2007; 8(1):16. [PubMed: 17555582]

15. Wood L, Egger M, Gluud L, Schulz K, Juni P, Altman D, Gluud C, Martin R, Wood A, Sterne J. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. Bmj. 336(7644):601–5.

16. Conover, WJ., editor. Practical Nonparametric Statistics. 3. John Wiley & Sons; 1999.

17. Rosner, B., editor. Fundamentals of Biostatistics. 6. Thomson, Brooks/Cole; 2006.

18. Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. J Clin Oncol. 1986 Jun; 4(6):942–51. [PubMed: 3711962]

19. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I: Medical. Stat Med. 1989 Apr; 8(4):441–54. [PubMed: 2727468]

20. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. Jama. 1995 Feb 1; 273(5):408–12. [PubMed: 7823387]

21. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. Bmj. 2001 Jul 7; 323(7303):42–6. [PubMed: 11440947]

22. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. Jama. 1996 Aug 28; 276(8):637–9. [PubMed: 8773637]

23. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Bm. 40:c332.

24. Godlee F. Publishing study protocols: making them visible will improve registration, reporting and recruitment. BMC News and Views. 2001; 2:4.

25. Horton R, Smith R. Time to register randomised trials. The case is now unanswerable. Bmj. 1999 Oct 2; 319(7214):865–6. [PubMed: 10506022]

26. Tonks A. Registering clinical trials. Bmj. 1999 Dec 11; 319(7224):1565–8. [PubMed: 10591727]

27. Simes RJ. Publication bias: the case for an international registry of clinical trials. J Clin Oncol. 1986; 4:1529–41. [PubMed: 3760920]
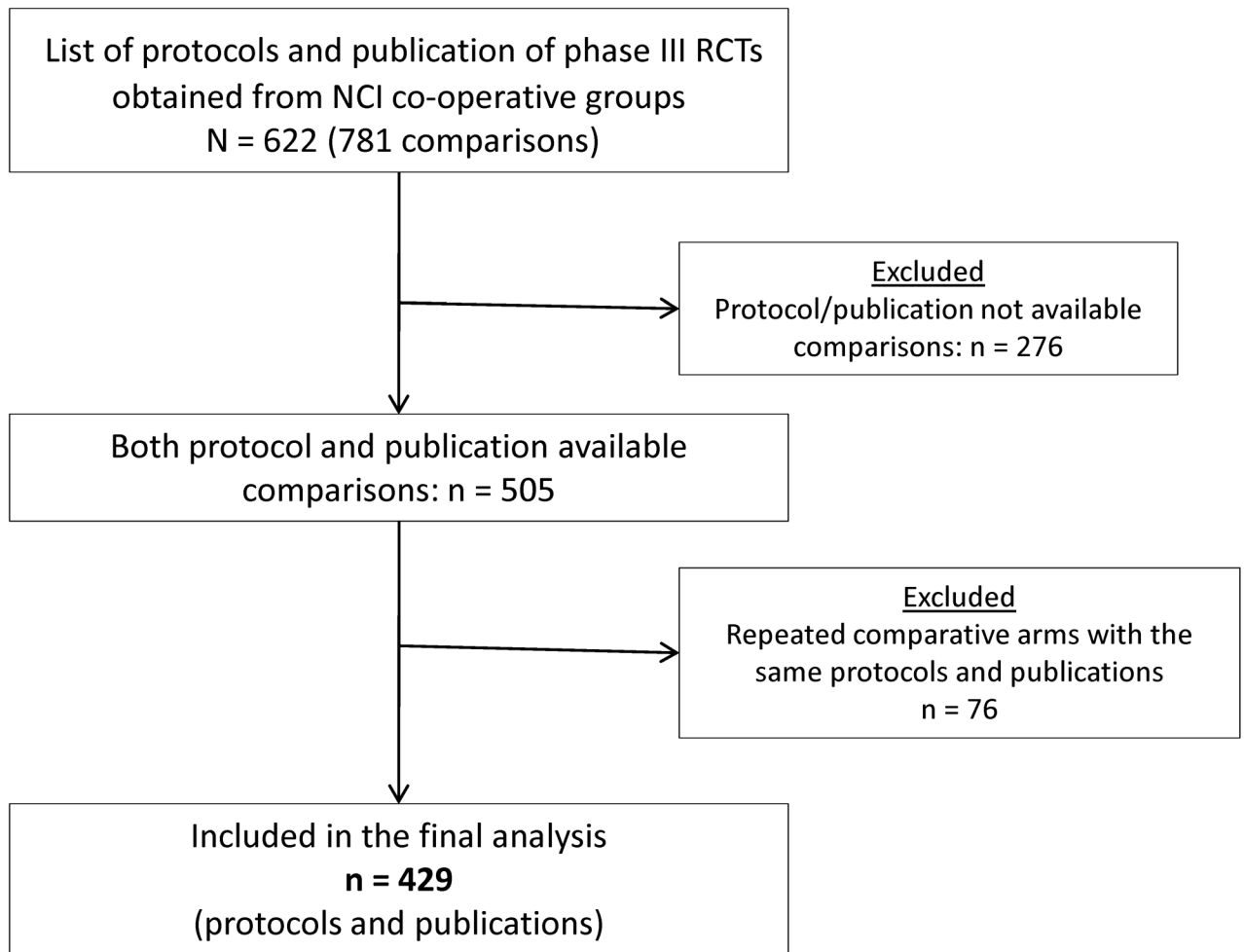
Figure 1.
Flow chart depicting the study selection process

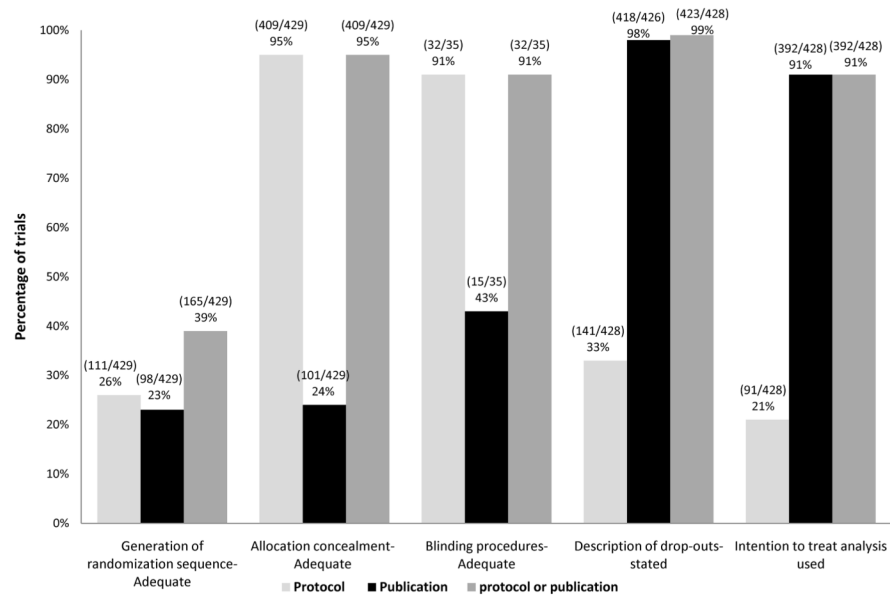**Figure 2.**
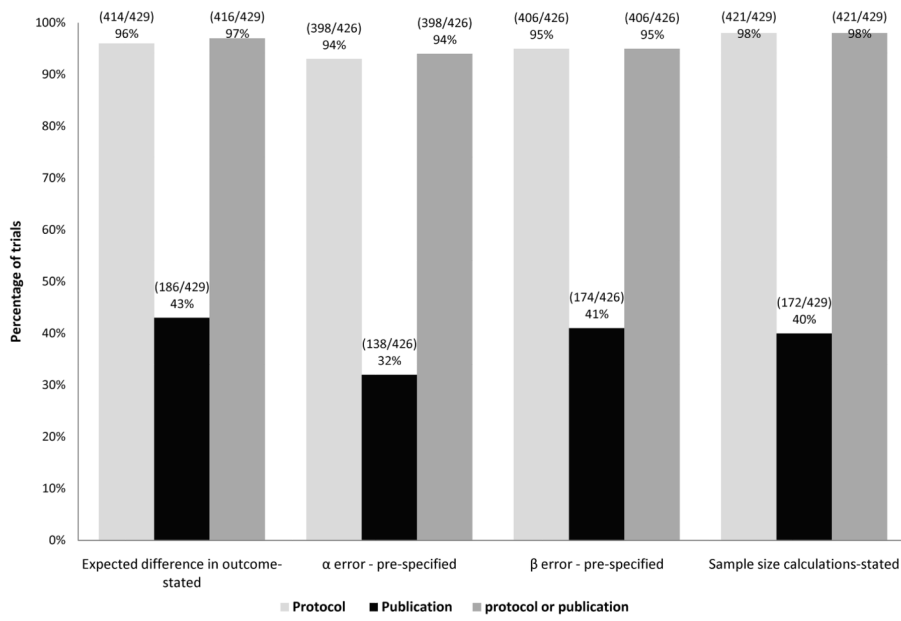Reporting of methodological quality domains associated with risk of bias in publication versus protocols.

Figure 3.
Reporting of methodological quality domains associated with risk of random error in publication versus protocols.

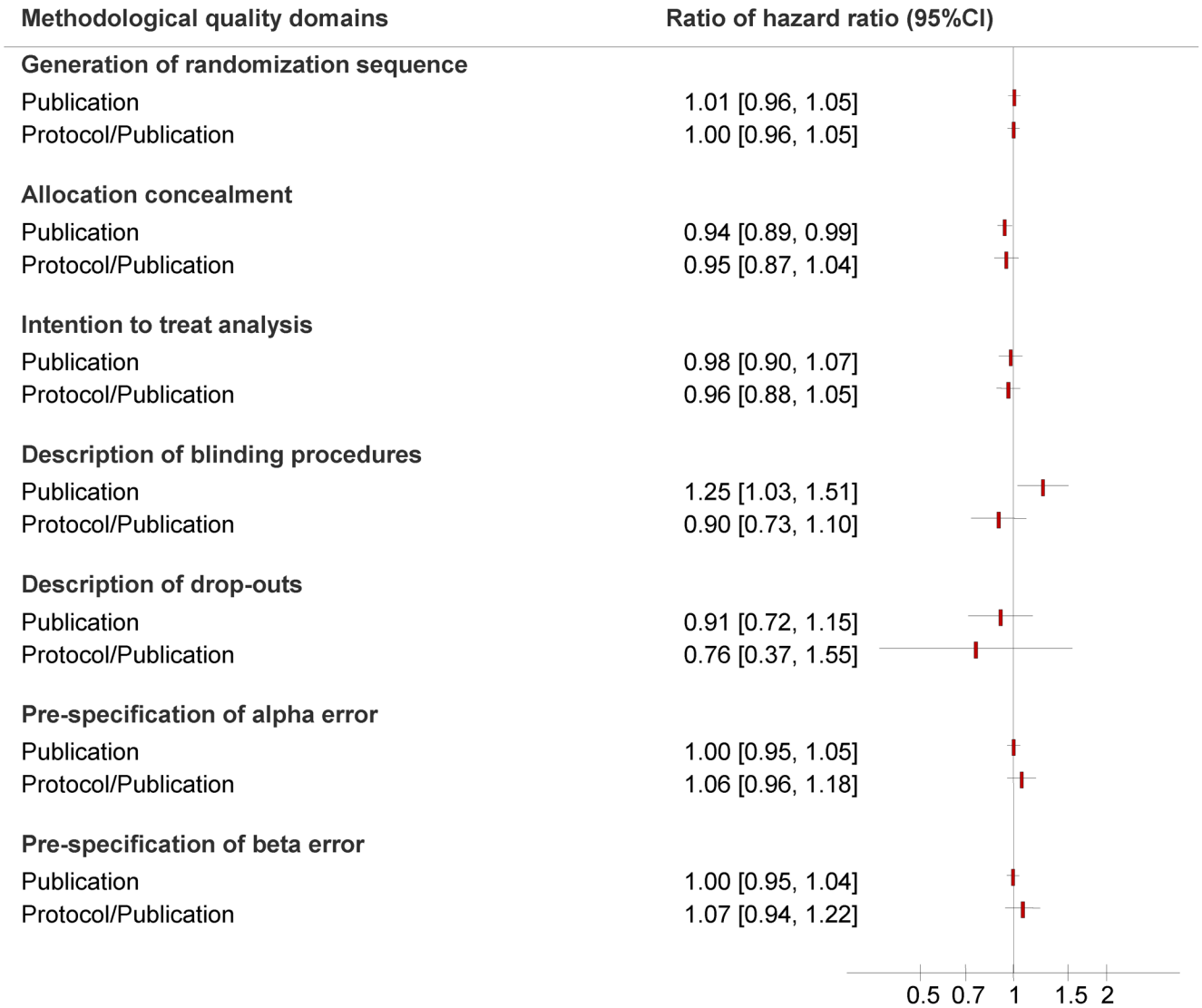| Methodological quality domains	Ratio of hazard ratio (95%CI)
Generation of randomization sequence	
Publication	1.01 [0.96, 1.05]
Protocol/Publication	1.00 [0.96, 1.05]
Allocation concealment	
Publication	0.94 [0.89, 0.99]
Protocol/Publication	0.95 [0.87, 1.04]
Intention to treat analysis	
Publication	0.98 [0.90, 1.07]
Protocol/Publication	0.96 [0.88, 1.05]
Description of blinding procedures	
Publication	1.25 [1.03, 1.51]
Protocol/Publication	0.90 [0.73, 1.10]
Description of drop-outs	
Publication	0.91 [0.72, 1.15]
Protocol/Publication	0.76 [0.37, 1.55]
Pre-specification of alpha error	
Publication	1.00 [0.95, 1.05]
Protocol/Publication	1.06 [0.96, 1.18]
Pre-specification of beta error	
Publication	1.00 [0.95, 1.04]
Protocol/Publication	1.07 [0.94, 1.22]

0.5 0.7 1   1.5 2

Figure 4.
Forest plot depicting ratio of hazard ratios for the outcome of overall survival (treatment effect) according to adequate versus inadequate reporting of methodological quality domains judged according to reporting in publications only versus protocol plus publications. Each rectangle represents the ratio of hazard ratios and the horizontal lines represent 95% confidence intervals (CI). A ratio of hazard ratios (RHR) < 1 implies that treatment effects were more beneficial in trials with adequate description of the methodological quality domain. A RHR of 1 implies that treatment effects were less beneficial in trials with adequate description of the methodological quality domain. A RHR of <1 along with associated CIs of <1 indicates a statistically significant association of methodological quality and treatment effect size.

Table 1

Association of reported methodological quality and sample size (publication data only)

Assessment of reported methodological quality parameters	N (%)	Median sample size (range)	Kruskal-Wallis test P-value
Allocation concealment			0.09
Adequate	101 (24)	215 (26–1387)	
Inadequate	328 (76)	242 (30–18882)	
Generation randomization sequence			
Adequate	98 (23)	262 (40–18882)	
Inadequate	331 (77)	236 (26–3561)	
Blinding procedures			0.05
Adequate	15 (43)	234 (47–1387)	
Inadequate	20 (57)	123 (45–18882)	
Description of drop outs			0.56
Adequate	418 (97)	241 (26–18882)	
Inadequate	8 (3)	192 (47–1695)	
Intention to treat analysis			0.16
Done and reported	392 (91)	352 (86–18882)	
Not done and not reported	36 (9)	216 (43–724)	
Pre specification of α error			0.14
Yes	138 (32)	252 (30–18882)	
No	290 (68)	237 (26–2233)	
Pre specification of β error			0.23
Yes	174 (41)	244 (30–18882)	
No	254 (59)	236 (26–3561)	
Choice of the comparator			< 0.001
Active Treatment	351 (82)	245 (26–18882)	
Placebo	30 (7)	132 (45–1387)	
Active Treatment + Placebo	3 (1)	260 (64–266)	
No Treatment	45 (10)	242 (30–4406)	

Table 2

Association of actual methodological quality and sample size (protocol/publication data)

Final assessment of methodological quality parameters (data extracted from either protocol/publication)	N (%)	Median sample size (range)	Kruskal-Wallis test P-value
Allocation concealment			0.8
Adequate	409 (95)	241 (26–18882)	
Inadequate	20 (5)	195 (83–1695)	
Generation randomization sequence			0.34
Adequate	165 (39)	243 (40–18882)	
Inadequate	264 (61)	237 (26–3561)	
Blinding procedures			0.88
Adequate	32 (91)	136 (45–18882)	
Inadequate	3 (9)	168 (48–331)	
Description of drop outs			0.16
Adequate	423 (99)	240 (26–18882)	
Inadequate	5 (1)	114 (47–322)	
Intention to treat analysis			0.16
Done and reported	392 (91)	242 (26–18882)	
Not done and not reported	36 (9)	196 (43–724)	
Pre specification of α error			0.06
Yes	398 (93)	243 (30–18882)	
No	28 (7)	163 (26–759)	
Pre specification of β error			0.06
Yes	406 (95)	243 (30–18882)	
No	20 (5)	163 (26–759)	