

BMJ Open Peer review comments on drug trials submitted to medical journals differ depending on sponsorship, results and acceptance: a retrospective cohort study

Marlies van Lent,¹ Joanna IntHout,² Henk Jan Out^{1,3}

To cite: van Lent M, IntHout J, Out HJ. Peer review comments on drug trials submitted to medical journals differ depending on sponsorship, results and acceptance: a retrospective cohort study. *BMJ Open* 2015;**5**:e007961. doi:10.1136/bmjopen-2015-007961

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-007961>).

Received 13 February 2015

Revised 29 April 2015

Accepted 22 May 2015



CrossMark

¹Department of Pharmacology—Toxicology, Clinical Research Centre Nijmegen, Radboud University Medical Center, Nijmegen, The Netherlands

²Department for Health Evidence, Radboud University Medical Center, Nijmegen, The Netherlands

³Teva Pharmaceuticals, Amsterdam, The Netherlands

Correspondence to

Marlies van Lent;
Marlies.vanLent@
radboudumc.nl

ABSTRACT

Objective: During peer review, submitted manuscripts are scrutinised by independent experts to assist journal editors in their decision-making and to help improve the quality of articles. In this retrospective cohort study, peer review comments for drug trials submitted to medical journals were analysed to investigate whether there is a relation between the content of these comments and sponsorship, direction of results and decision about acceptance.

Design/setting: Descriptive content analysis of reviewer comments made on manuscripts on drug trials submitted to eight medical journals (January 2010–April 2012). For each manuscript, the number of reviewers, decision about acceptance, sponsorship and direction of results were extracted. Reviewer comments were classified using a predefined checklist.

Results: Reviewer reports for 246 manuscripts were assessed. Industry-sponsored trials were more likely to receive comments about lack of novelty (8.9%) than industry-supported (2.5%) and non-industry trials (6.1%, overall $p=0.038$). Non-industry trials more often received comments about poor experimental design (69.7%) than industry-supported (58.8%) and industry-sponsored trials (52.9%, overall $p=0.019$). Non-industry trials were also more likely to receive comments regarding inappropriate statistical analyses (28.4%) than industry-supported (23.5%) and industry-sponsored trials (15.1%, overall $p=0.006$). Manuscripts with negative results were more likely to receive comments about inappropriate conclusions (29.3%) than those with positive results (18.9%, $p=0.010$). Rejected manuscripts had more often received comments on the research question not being clinically relevant (7.8%) than accepted manuscripts (1.6%, $p=0.002$), and also on lack of novelty (8.3% vs 2.6%, $p=0.008$) and poor experimental design (68.6% vs 50.5%, $p<0.001$).

Conclusions: Reviewers identified fewer shortcomings regarding design and statistical analyses in industry-related trials, but commented more often on a lack of novelty in industry-sponsored trials. Negative trial results did not significantly influence the nature of comments other than appropriateness of the conclusion. Manuscript acceptance was primarily

Strengths and limitations of this study

- Analysis of real life peer review comments on submitted manuscripts in relation to sponsorship, direction of results and decision about acceptance.
- Inclusion of manuscripts submitted to a general medical journal and specialty journals across different medical specialties.
- Comprehensiveness of classification checklist assessed in two training sessions. Reviewer comments for 20% of included manuscripts scored by two raters and good level of inter-rater agreement.
- Focus on reviewer comments made during initial peer review of articles; additional comments may have been raised during reviews of revised versions.
- Reviewer comments may not inherently provide an objective reflection of the quality of articles, as there is evidence in the literature on the defects of peer review.

related to the research question and methodological robustness of studies.

INTRODUCTION

At peer-reviewed medical journals, submitted articles are sent out for external peer review if they are considered to be potentially suitable for publication. During peer review, manuscripts are scrutinised by independent experts or peers in the same field, to assist editors in their decision-making and to help improve the quality of submitted articles.^{1 2} The peer review process has been investigated to a limited extent. Some studies addressed the effects of blinding and training of reviewers, the detection rate of deliberately introduced errors by reviewers, and the impact of peer review on the quality of published articles.^{3–8}

Few studies have systematically analysed the content of reviewer comments on submitted articles. Bordage⁹ studied the reasons given by reviewers for rejection of manuscripts submitted for publication in conference proceedings on research in medical education. Inappropriate statistics and overinterpretation of results were commonly reported.⁹ Turcotte *et al*¹⁰ analysed reviewer comments for manuscripts submitted to the Canadian Journal of Anesthesia, and found that lack of originality, inadequate experimental design and inappropriate conclusions were the main determinants of an article's fate. Hopewell *et al*⁷ focused on reviewer comments on the reporting of methodological items in randomised trials submitted to open peer review journals. The type of changes requested by reviewers included addition or clarification of randomisation, blinding, and sample size, and toning down of conclusions to reflect the results.⁷

The content of reviewer comments may be related to the direction of results and sponsorship. Emerson *et al*¹¹ compared reviewer reports for a fabricated manuscript reporting positive results and an otherwise identical manuscript reporting no effect. Reviewers detected more errors in the no-difference version and awarded higher scores to the methodology section of the positive manuscript, although the methods sections were identical.¹¹ Emerson *et al*¹¹ showed that the positive article was more often recommended for publication than the no-difference version, although this observation could not be confirmed by others.¹² None of the previous studies on peer review compared reviewer comments according to whether reported trials were sponsored by pharmaceutical companies or non-profit organisations.

Analysis of reviewer comments provides more insight into the shortcomings of drug trials that are submitted to medical journals, both from the perspective of the design and conduct of trials and the reporting quality in articles. It would be interesting to determine whether the occurrence of specific shortcomings in manuscripts is affected by sponsorship and results being either positive or negative. In the current study, we performed a descriptive content analysis of peer review comments made on manuscripts on drug trials submitted to eight medical journals to investigate the relationship between the content of comments and sponsorship, direction of results, and decision about acceptance, using a previously reported cohort.¹³

METHODS

Journal and manuscript selection

We included manuscripts submitted from January 2010 through April 2012 to one general medical journal (BMJ) and seven specialty journals (Annals of the Rheumatic Diseases, British Journal of Ophthalmology, Gut, Heart, Thorax (all from the BMJ Group), Diabetologia, and Journal of Hepatology). We selected randomised controlled trials, in which at least one study

arm assessed the efficacy or safety of a drug and a statistical test was used to evaluate treatment effects. This cohort has been described in detail previously.¹³ This study was limited to manuscripts that were sent out for external peer review.

Data extraction manuscripts

For each manuscript, the decision about acceptance and reviewer reports were extracted from submission systems or provided by journals. Manuscripts were either rejected after review or accepted for publication. We determined the number of reviewers that evaluated the first submitted version of each article. Manuscripts could be evaluated during multiple rounds of review before a final decision was made, but we focused on comments that were made during initial peer review of articles. Reviews of revised manuscripts were excluded. Information on sponsorship and the direction of results was previously extracted from manuscripts and classified according to predefined criteria.¹⁴ Reviewers were aware of the sponsorship of trials. In short, trials were classified as non-industry, industry-supported or industry-sponsored trials. For non-industry trials, no associations with pharmaceutical companies were reported. Studies reporting donation of study medication by a manufacturer, studies stating receipt of financial support from a pharmaceutical company, and studies with industry-affiliated authors were classified as industry-supported trials. For industry-sponsored trials, a pharmaceutical company was explicitly described as study sponsor, or the company funding the trial participated in the design, data collection, analysis and/or preparation of the manuscript. Trial results were scored as positive if results reported for the primary end point were statistically significant ($p < 0.05$ or 95% CI for difference excluding 0 or 95% CI for ratio excluding 1) and supported the efficacy of the test drug, and negative if they did not. Results of non-inferiority trials were classified as positive if treatments were equivalent. Safety trials were classified as positive if the test drug was as safe as or safer than control.

Classification of reviewer comments

A validated instrument for classification of reviewer comments does not exist. Included journals did not provide standardised forms to reviewers, but general guidance for peer review was available on each journals' website. Based on this guidance^{15–17} and previous research on peer review,^{9 10 18–20} a classification checklist for negative reviewer comments was composed. In two consecutive training sessions, reviewer comments for 10 randomly selected manuscripts were independently classified by two raters (MvL and HJO) in each session, to assess the consistency between raters and check on the comprehensiveness of the checklist. Both after the first and second training session, disparities in the interpretation of comments were discussed and the checklist was revised accordingly. The final version of the checklist

(see online supplementary table S1) was then tested on reviewer comments for a random sample of 30 manuscripts that were independently classified by the two raters. Assuming an inter-rater agreement of at least 80% for each type of comment with the final checklist, 30 manuscripts were sufficient to estimate the agreement with a precision (SE) of at most 7%. If the inter-rater agreement during this test was considered sufficiently high, a single reviewer (MvL) could continue with the rating process. After the classification of reviewer comments for these 30 manuscripts, we calculated the percentage of agreement between raters for each type of comment in the checklist. κ Statistics were not considered suitable as some types of comments were rarely scored and resulting κ values would be inaccurate.²¹ For these 30 manuscripts, classification discrepancies were resolved by consensus between raters if the agreement for a comment was <85%. For the other types of comments, the score assigned by the rater who subsequently classified all comments for the other manuscripts (MvL) was decisive. Overall, reviewer comments for 50 manuscripts were scored by two raters in this study, which was equivalent to 20% (50 of 246) of the total number of included manuscripts. For each manuscript, a reviewer could have several remarks related to one type of comment. However, each type of comment was scored maximally once per reviewer.

Statistical analysis

Descriptive statistics were used to describe included manuscripts (data presented as frequencies and percentages). The relationship between each type of comment and sponsorship, direction of results and decision about acceptance was analysed using a generalised linear mixed model based on generalised estimating equations (GEE) with a binary distribution for the dependent variable and an identity link. In this model, the comment score of a reviewer (comment vs no comment) was used as the dependent variable. We included sponsor type, results or decision about acceptance as fixed variable in the model, and—if possible (if the model converged)—journal, to control for the journal to which a manuscript was submitted. The unique identification number that manuscripts received from a journal was included as cluster variable (random effect). Most manuscripts were reviewed by several reviewers. The model estimates the percentage of reviewers that will comment on a manuscript, depending on sponsor type, results or decision about acceptance ('mean percentage of comments on a manuscript'). If a lower limit of the resulting CI was negative, it was truncated to 0. The number of different types of comments per manuscript was compared by sponsorship, results or decision about acceptance using univariate analysis of variance. We controlled for the number of reviewers per manuscript by including this as a covariate in the model. Two-sided $p < 0.05$ was considered statistically significant. p Values were not adjusted for multiple comparisons. Statistical analyses were

performed using SPSS software (V.20) and SAS for Windows (V.9.2, SAS Institute Inc).

Ethics

To assure confidentiality of manuscripts and reviewer reports, confidentiality agreements were signed by the authors before gaining access to the data. As standard editorial and peer review processes were unchanged, authors and reviewers were not informed about this study. Research ethics committee (REC) approval was not required as this study involved no human participants.

RESULTS

From January 2010 through April 2012, 472 manuscripts on drug RCTs were submitted to eight journals, of which 250 articles (53.0%) were externally reviewed. For 246 manuscripts, reviewer comments for authors were available. Of these 246, 96 (39.0%) were accepted for publication (table 1). Eighty-nine (36.2%) were non-industry trials, while 78 (31.7%) were industry-supported and 79 (32.1%) were industry-sponsored trials. Most articles reported positive results ($N=150$, 61.0%). The number of reviewers for the first submitted version of an article ranged from 1 to 5. In total, 575 reviewer reports were evaluated.

Overall, the level of inter-rater agreement for the final version of the classification checklist was good. For all

Table 1 Characteristics of included manuscripts

	Manuscripts, n (%)
Total	246 (100)
Journal	
<i>BMJ</i>	39 (15.9)
<i>Annals of the Rheumatic Diseases</i>	44 (17.9)
<i>British Journal of Ophthalmology</i>	12 (4.9)
<i>Gut</i>	31 (12.6)
<i>Heart</i>	8 (3.3)
<i>Thorax</i>	23 (9.3)
<i>Diabetologia</i>	51 (20.7)
<i>Journal of Hepatology</i>	38 (15.4)
Decision about acceptance	
Rejected after peer review	150 (61.0)
Accepted for publication	96 (39.0)
Sponsor type	
Non-industry	89 (36.2)
Industry-supported	78 (31.7)
Industry-sponsored	79 (32.1)
Trial results	
Positive results	150 (61.0)
Negative results	96 (39.0)
Number of reviewers per manuscript	
1	10 (4.1)
2	160 (65.0)
3	61 (24.8)
4	13 (5.3)
5	2 (0.8)

types of comments, the agreement between raters was close to or higher than 80%. For 20 of 26 items, there was >85% agreement (see online supplementary table S2).

Overall, the types of comments that were most frequently reported by reviewers included poor experimental design (range of point estimators in tables 2 and 3, and 4; 50.5–69.7%), inadequately reported methods (50.8–60.5%), incomplete study outcome data (58.7–68.2%), inadequate discussion of the meaning of results

(44.2–56.1%), poor writing (34.7–42.8%) and inaccurate tables or figures (35.1–44.1%). In table 2, the mean percentage of comments on a manuscript is compared by sponsor type. For several types of comments, there was a relation between sponsorship and the mean percentage of comments. The percentage of comments regarding a lack of novelty was significantly associated with sponsorship ($p=0.038$); industry-sponsored trials were more likely to receive this comment (8.9%) than industry-supported (2.5%) and non-industry trials (6.1%).

Table 2 Distribution of reviewer comments—by sponsor type

Type of comment	Mean percentage of comments on a manuscript (95% CI)*			p Value
	Non-industry (N=89)	Industry-supported (N=78)	Industry-sponsored (N=79)	
Importance				
1. Research question not clinically relevant	6.3 (2.4 to 10.2)	6.1 (2.2 to 10.1)	3.3 (0.3 to 6.3)	0.372
Originality				
2. Lack of novelty	6.1 (2.6 to 9.7)	2.5 (0.1 to 4.9)	8.9 (4.1 to 13.7)	0.038
Background and rationale				
3. Incorrect background information	20.4 (15.2 to 25.5)	18.4 (12.2 to 24.6)	18.8 (12.5 to 25.2)	0.877
4. Poor justification for conducting study	1.5 (0.0 to 3.1)	2.8 (0.4 to 5.1)	6.3 (2.4 to 10.1)	0.081
Methods				
5. Poor experimental design	69.7 (63.1 to 76.3)	58.8 (50.2 to 67.4)	52.9 (43.9 to 61.9)	0.019
6. Methods inadequately reported	60.5 (53.9 to 67.1)	54.7 (46.7 to 62.7)	50.8 (42.4 to 59.2)	0.209
7. Statistical analysis methods inappropriate	28.4 (22.3 to 34.6)	23.5 (16.4 to 30.5)	15.1 (10.1 to 20.2)	0.006
Results				
8. Study outcome data incomplete	65.9 (59.4 to 72.4)	68.0 (59.7 to 76.4)	58.7 (50.6 to 66.8)	0.215
9. Flow of participants through study unclear	7.7 (3.8 to 11.6)	7.8 (3.3 to 12.4)	4.6 (1.8 to 7.4)	0.323
Discussion and conclusion				
10. Meaning results inadequately discussed	44.2 (36.6 to 51.9)	46.7 (38.5 to 54.9)	56.1 (47.5 to 64.7)	0.090
11. Study insufficiently related to literature	15.2 (10.4 to 20.0)	15.5 (8.3 to 22.6)	8.7 (4.2 to 13.3)	0.180
12. Limitations not sufficiently discussed	17.2 (11.6 to 22.8)	19.9 (14.7 to 25.1)	13.8 (8.3 to 19.3)	0.223
13. Conclusions inappropriate	24.2 (17.7 to 30.6)	23.0 (16.0 to 30.1)	20.0 (13.2 to 26.8)	0.652
Abstract				
14. Abstract does not correctly reflect paper	16.2 (11.5 to 20.9)	17.1 (11.8 to 22.4)	14.4 (8.8 to 19.9)	0.768
15. Discrepancies between the abstract and the main text	2.0 (0.1 to 3.9)	0.6 (0.0 to 1.6)	1.0 (0.0 to 2.5)	0.443
References				
16. References missing/irrelevant references used	11.4 (7.0 to 15.8)	12.0 (6.6 to 17.4)	11.5 (6.1 to 16.8)	0.985
17. Errors in reference citation	1.5 (0.0 to 3.1)	1.7 (0.0 to 3.5)	4.6 (1.7 to 7.5)	0.159
Presentation				
18. Title not representative of study	5.0 (1.7 to 8.2)	8.1 (3.3 to 12.8)	1.5 (0.0 to 3.1)	0.012
19. Poor writing	42.8 (35.5 to 50.2)	35.4 (27.9 to 42.9)	34.7 (26.8 to 42.6)	0.258
20. Inaccurate tables or figures	37.0 (30.9 to 43.1)	44.1 (36.5 to 51.7)	37.2 (29.5 to 45.0)	0.306
Ethics				
21. Ethics committee approval not clear	2.0 (0.1 to 3.9)	1.7 (0.0 to 3.5)	2.1 (0.1 to 4.1)	0.951
22. Other ethical issues related to study	3.1 (0.0 to 6.3)	4.1 (0.7 to 7.5)	2.0 (0.1 to 4.0)	0.555
Trial registration, protocol, CONSORT				
23. Registration/protocol/CONSORT missing	2.5 (0.4 to 4.6)	2.8 (0.4 to 5.1)	2.6 (0.4 to 4.8)	0.984
24. Deviations from registry or protocol	1.4 (0.0 to 3.3)	1.8 (0.0 to 3.8)	1.7 (0.0 to 3.6)	0.961
COI				
25. Bias by author COIs/contribution funder unclear	2.5 (0.4 to 4.6)	2.3 (0.1 to 4.4)	3.6 (1.1 to 6.2)	0.707
26. Systematic bias or spin in favour of sponsor	0.0 (0.0 to 1.1)	0.5 (0.0 to 1.8)	1.7 (0.2 to 3.1)	0.139

*The mean percentage of comments on a manuscript is controlled for the journal to which a manuscript was submitted. COI, conflicts of interest; N, number of submitted manuscripts.

Table 3 Distribution of reviewer comments—by direction of trial results

Type of comment	Mean percentage of comments on a manuscript (95% CI)*		
	Negative results (N=96)	Positive results (N=150)	p Value
Importance			
1. Research question not clinically relevant	4.3 (1.1 to 7.5)	5.9 (3.1 to 8.7)	0.469
Originality			
2. Lack of novelty	3.5 (0.6 to 6.3)	7.4 (4.4 to 10.4)	0.066
Background and rationale			
3. Incorrect background information	20.3 (14.9 to 25.8)	18.9 (14.5 to 23.3)	0.670
4. Poor justification for conducting study	4.2 (1.3 to 7.2)	3.0 (1.2 to 4.9)	0.497
Methods			
5. Poor experimental design	64.9 (56.9 to 73.0)	60.0 (54.5 to 65.5)	0.285
6. Methods inadequately reported	55.4 (48.2 to 62.6)	56.3 (50.7 to 61.9)	0.844
7. Statistical analysis methods inappropriate	23.7 (17.3 to 30.1)	20.5 (15.7 to 25.3)	0.453
Results			
8. Study outcome data incomplete	61.4 (53.5 to 69.2)	66.4 (61.0 to 71.8)	0.266
9. Flow of participants through study unclear	5.5 (2.3 to 8.8)	7.4 (4.5 to 10.3)	0.411
Discussion and conclusion			
10. Meaning results inadequately discussed	47.8 (40.0 to 55.7)	48.5 (42.1 to 54.8)	0.889
11. Study insufficiently related to literature	12.6 (7.7 to 17.5)	13.7 (9.7 to 17.7)	0.748
12. Limitations not sufficiently discussed	14.4 (9.4 to 19.3)	18.4 (14.3 to 22.4)	0.203
13. Conclusions inappropriate	29.3 (22.6 to 36.0)	18.9 (14.0 to 23.9)	0.010
Abstract			
14. Abstract does not correctly reflect paper	16.4 (11.4 to 21.4)	15.5 (11.8 to 19.2)	0.790
15. Discrepancies between the abstract and the main text	1.9 (0.1 to 3.6)	0.8 (0.0 to 1.8)	0.320
References			
16. References missing/irrelevant references used	14.9 (9.8 to 20.0)	8.6 (5.8 to 11.5)	0.079
17. Errors in reference citation	2.3 (0.3 to 4.3)	2.8 (1.1 to 4.4)	0.731
Presentation			
18. Title not representative of study	6.7 (2.5 to 10.9)	3.6 (1.7 to 5.5)	0.191
19. Poor writing	38.4 (31.4 to 45.4)	38.4 (32.5 to 44.3)	0.997
20. Inaccurate tables or figures	35.1 (28.6 to 41.6)	41.4 (36.1 to 46.7)	0.156
Ethics			
21. Ethics committee approval not clear	1.4 (0.0 to 3.0)	2.2 (0.7 to 3.7)	0.463
22. Other ethical issues related to study	2.0 (0.1 to 3.9)	3.8 (1.4 to 6.2)	0.267
Trial registration, protocol, CONSORT			
23. Registration/protocol/CONSORT missing	1.9 (0.1 to 3.6)	3.0 (1.3 to 4.8)	0.352
24. Deviations from registry or protocol	1.8 (0.0 to 4.0)	1.4 (0.2 to 2.7)	0.758
COI			
25. Bias by author COIs/contribution funder unclear	1.9 (0.1 to 3.6)	3.3 (1.5 to 5.1)	0.254
26. Systematic bias or spin in favour of sponsor	0.0 (0.0 to 1.1)	1.0 (0.0 to 2.0)	0.120

*The mean percentage of comments on a manuscript is controlled for the journal to which a manuscript was submitted. COI, conflicts of interest; N, number of submitted manuscripts.

The percentage of comments regarding poor experimental design was also associated with sponsorship ($p=0.019$); non-industry trials more often received this comment (69.7%) than industry-supported (58.8%) and industry-sponsored trials (52.9%). Furthermore, the percentage of comments about inappropriate statistical analysis methods was associated with sponsorship ($p=0.006$); non-industry trials were more likely to receive this comment (28.4%) than industry-supported (23.5%) and industry-sponsored trials (15.1%). The percentage of comments regarding the article title not being representative of the study was also associated with sponsorship ($p=0.012$); industry-supported trials more often received this comment (8.1%) than non-industry (5.0%) and industry-sponsored trials (1.5%).

In table 3, the mean percentage of comments on a manuscript is compared by the direction of trial results. For most types of comments, there was no significant difference according to whether manuscripts reported positive or negative results. However, the percentage of comments regarding inappropriate conclusions was higher for articles with negative trial results (29.3%) than for articles with positive results (18.9%, $p=0.010$).

Table 4 shows the mean percentage of comments on a manuscript according to the decision about acceptance. The percentage of comments about the research question not being clinically relevant was higher among rejected manuscripts (7.8%) than accepted manuscripts (1.6%, $p=0.002$). Rejected manuscripts were more likely to receive comments regarding a lack of novelty (8.3%)

Table 4 Distribution of reviewer comments—by decision about acceptance

Type of comment	Mean percentage of comments on a manuscript (95% CI)*		
	Rejected (N=150)	Accepted (N=96)	p Value
Importance			
1. Research question not clinically relevant	7.8 (4.5 to 11.1)	1.6 (0.0 to 3.5)	0.002
Originality			
2. Lack of novelty	8.3 (4.9 to 11.6)	2.6 (0.6 to 4.7)	0.008
Background and rationale			
3. Incorrect background information	20.3 (15.8 to 24.8)	18.0 (12.7 to 23.3)	0.491
4. Poor justification for conducting study	2.4 (0.8 to 4.1)	4.9 (1.9 to 7.9)	0.168
Methods			
5. Poor experimental design	68.6 (63.3 to 73.9)	50.5 (42.8 to 58.2)	<0.001
6. Methods inadequately reported	57.7 (51.8 to 63.6)	52.7 (45.5 to 59.9)	0.306
7. Statistical analysis methods inappropriate	21.8 (16.7 to 27.0)	20.8 (15.1 to 26.4)	0.793
Results			
8. Study outcome data incomplete	62.4 (56.4 to 68.5)	68.2 (61.3 to 75.1)	0.178
9. Flow of participants through study unclear	8.2 (4.8 to 11.5)	4.8 (2.3 to 7.2)	0.110
Discussion and conclusion			
10. Meaning results inadequately discussed	45.5 (39.1 to 51.9)	53.1 (45.6 to 60.6)	0.092
11. Study insufficiently related to literature	10.5 (7.1 to 14.0)	18.6 (13.0 to 24.2)	0.041
12. Limitations not sufficiently discussed	16.0 (12.3 to 19.8)	18.3 (12.8 to 23.8)	0.454
13. Conclusions inappropriate	23.1 (17.9 to 28.3)	21.8 (15.4 to 28.2)	0.755
Abstract			
14. Abstract does not correctly reflect paper	15.7 (12.0 to 19.4)	16.1 (11.2 to 21.1)	0.891
15. Discrepancies between the abstract and the main text	1.5 (0.2 to 2.8)	0.8 (0.0 to 1.9)	0.419
References			
16. References missing/irrelevant references used	11.6 (8.0 to 15.2)	11.6 (6.6 to 16.7)	0.990
17. Errors in reference citation	2.7 (1.0 to 4.4)	2.5 (0.6 to 4.3)	0.843
Presentation			
18. Title not representative of study	6.2 (3.1 to 9.2)	2.9 (0.8 to 4.9)	0.080
19. Poor writing	37.5 (31.6 to 43.3)	40.0 (32.7 to 47.4)	0.555
20. Inaccurate tables or figures	36.7 (31.5 to 41.9)	43.3 (36.3 to 50.2)	0.121
Ethics			
21. Ethics committee approval not clear	1.8 (0.4 to 3.2)	2.1 (0.3 to 3.8)	0.839
22. Other ethical issues related to study	4.1 (1.5 to 6.7)	1.6 (0.1 to 3.2)	0.115
Trial registration, protocol, CONSORT			
23. Registration/protocol/CONSORT missing	3.0 (1.2 to 4.9)	2.0 (0.3 to 3.8)	0.435
24. Deviations from registry or protocol	2.2 (0.6 to 3.9)	0.8 (0.0 to 2.2)	0.217
COI			
25. Bias by author COIs/contribution funder unclear	3.7 (1.7 to 5.6)	1.6 (0.1 to 3.2)	0.116
26. Systematic bias or spin in favour of sponsor	1.1 (0.1 to 2.1)	0.0 (0.0 to 0.9)	0.054

*The mean percentage of comments on a manuscript is controlled for the journal to which a manuscript was submitted.
COI, conflicts of interest; N, number of submitted manuscripts.

than accepted manuscripts (2.6%, $p=0.008$). In addition, the percentage of comments about poor experimental design was higher for rejected manuscripts (68.6%) than for those that were accepted (50.5%, $p<0.001$). Reviewers more often reported that a study was insufficiently related to the literature among manuscripts that were accepted (18.6%) compared to those that were rejected (10.5%, $p=0.041$).

In table 5, the number of different types of comments per manuscript is shown, which was adjusted for the number of reviewers per manuscript. Overall, reviewers reported a mean number of 7.8 different types of comments per manuscript (range, 1–15 types of comments). The number of types of comments per manuscript was not associated with the direction of results or the

decision about acceptance of manuscripts. There was a significant relation between sponsorship and the number of different types of comments per manuscript ($p=0.035$); non-industry trials on average received more types of comments per manuscript (8.2) than industry-sponsored trials (7.2).

DISCUSSION

This is the first study in which real life peer review comments made on submitted manuscripts were compared according to sponsorship, direction of results and decision about acceptance. Previous studies have been limited to experiments with fictitious manuscripts.^{11 12} The most frequently reported comments by reviewers

Table 5 Number of different types of comments per article

	Mean number of types of comments (95% CI)*	p Value
Sponsor type		0.035
Non-industry	8.2 (7.6 to 8.8)†	
Industry-supported	8.0 (7.4 to 8.6)	
Industry-sponsored	7.2 (6.6 to 7.8)†	
Trial results		0.794
Positive	7.8 (7.3 to 8.2)	
Negative	7.9 (7.3 to 8.4)	
Decision about acceptance		0.145
Rejected	8.0 (7.6 to 8.5)	
Accepted	7.5 (6.9 to 8.0)	

*The mean number of types of comments per manuscript is controlled for the number of reviewers per manuscript.

†The p value for the mean difference between non-industry and industry-sponsored trials is <0.05.

included poor experimental design, inadequately reported methods, incomplete study outcome data, inadequate discussion of the meaning of results, poor writing, and inaccurate tables or figures, which is in line with findings of previous studies.^{7 9 10 18} Reviewers rarely reported on ethics, trial registration, or conflicts of interest, as was expected from prior research.^{10 19 20}

Submitted manuscripts on industry-sponsored trials more often received comments regarding a lack of novelty compared to industry-supported and non-industry trials. However, we found no significant difference according to sponsor type for comments on the clinical relevance of research questions. It has been argued in literature that studies by pharmaceutical companies may be less innovative than non-industry studies. Drug companies may more often focus on late-stage drug development and producing variations of drugs already on the market, while academia may be more likely to perform creative, early-stage clinical research.^{22 23} Interestingly, industry-supported trials were least often criticised by reviewers for lack of novelty. This may suggest that collaboration between academia and the pharmaceutical industry could potentially lead to more innovative clinical studies.

Non-industry trials were more likely to receive comments regarding poor experimental design and inappropriate statistical analysis methods than industry-supported and industry-sponsored trials. In addition, non-industry trials received significantly more different types of comments per manuscript than industry-sponsored trials. Prior research based on published articles showed that the methodological quality of trials funded by pharmaceutical companies was equal to or tended to be higher than that of non-industry trials.^{24–27} Previously, we studied the shortcomings of protocols of drug trials that were submitted for approval to RECs.²⁸ Based on the comments raised during REC review, we found that non-industry trials more often had

shortcomings regarding methodology and statistical analyses than industry-sponsored trials,²⁸ which is in line with findings of the current study.

Manuscripts with negative results were more likely to receive comments regarding overinterpretation or inappropriate conclusions in relation to results than manuscripts with positive results. The number of types of comments per manuscript was not associated with the direction of results. Evidence of inconsistencies between results and the interpretation of findings has previously been shown for published articles, especially among those with negative results.^{29 30} Authors may shape the impression of results in articles, that is, to add 'spin' to reports. Spin includes the use of specific reporting strategies to highlight that the experimental treatment is effective, despite non-significant results for the primary outcome, or to distract readers from non-significant results. This distorts the interpretation of results and misleads readers.^{30 31}

Rejected manuscripts had more often received comments on the research question not being clinically relevant, lack of novelty and poor experimental design than accepted papers. The number of types of comments per manuscript was not associated with decision about acceptance though. Although we found significant differences between comments for articles that were eventually rejected or accepted, there are many reasons why papers can get rejected beyond what is in reviewer reports for the initial submitted version of manuscripts. Moreover, editorial processes and the amount of weight put on reviewer comments when making publication decisions can be very variable across journals. Papers that reviewers are positive about may be rejected, while others are published despite of negative reviewer comments. As manuscript review by journals is a complicated and multistage process, it is difficult to determine the exact influence of reviewer comments in editorial decision-making.

This study is strengthened by the inclusion of manuscripts submitted to a general medical journal and specialty journals across different medical specialties. The studies by Bordage⁹ and Turcotte *et al*¹⁰ were limited to articles on research in medical education or anaesthesia, which reduced the generalisability of their findings. Hopewell *et al* included open peer review journals where reviewer comments are included alongside published articles. Reviewers may more often provide rather uncritical comments when reviewing for such journals, as they may fear reprisals for criticising other researchers' work openly.^{7 32} In this study, we assessed the comprehensiveness of the classification checklist in two training sessions. Reviewer comments for 20% of the included manuscripts were scored by two raters and the level of inter-rater agreement was good. In previous studies, the classification of reviewer comments was completely conducted by a single author.^{9 10}

This study has some limitations. We focused on peer review comments for the first submitted version of articles. Some journals may send revised versions to new

reviewers or back to the same reviewers. By focusing on reviews of initial versions, new comments raised during reviews of revisions may have been missed. However, initial reviewer reports often contain the most extensive comments and provide adequate information to compare reviewer comments according to sponsorship, direction of results and decision about acceptance. We have not assessed whether shortcomings that were detected by reviewers were corrected in revised manuscripts. Hopewell *et al*⁷ found that most authors complied with requests by reviewers in their revised version, but this was beyond the scope of this study. In addition, we included a sample of manuscripts describing drug RCTs and our results may therefore not be generalisable to other study designs or RCTs with other interventions.

Although peer review is generally assumed to raise the quality of submitted papers and to provide a mechanism for rational and fair editorial decision-making,³³ reviewer comments may not automatically provide an objective reflection of the quality of articles. While the evidence on the effectiveness of peer review is limited,^{2 6 33} there is considerable evidence on its defects.^{2 34} In studies where major errors were inserted into papers that were subsequently sent to reviewers, none of the reviewers spotted all of the errors.^{4 35} In addition, it has been suggested that peer review is a subjective and, therefore, inconsistent process.³⁴ Agreement between reviewers in their recommendations for manuscripts may be low.³⁶ Nevertheless, peer review is seen by researchers as important and essential for scientific communication and as the best alternative currently available.^{34 37}

In conclusion, peer reviewers identified fewer shortcomings regarding design and statistical analyses in industry-related trials, but commented more often on a lack of novelty in industry-sponsored trials. Negative trial results did not significantly influence the nature of comments other than appropriateness of the conclusion. Manuscript acceptance was primarily related to the research question and methodological robustness of the study. As some of the manuscripts' shortcomings represent fundamental methodological weaknesses, better training on trial design and analysis may be appropriate, especially for non-industry trials. Other errors are more just omissions, including frequently reported shortcomings such as inadequate reporting of methods and incomplete reporting of study outcome data. These fixable errors can be avoided if authors pay more attention to reporting quality in manuscripts.

Acknowledgements The authors thank BMJ, Annals of the Rheumatic Diseases, British Journal of Ophthalmology, Gut, Heart, Thorax, Diabetologia, and Journal of Hepatology for participating in this study. The authors thank Sara Schroter (BMJ) for her suggestions regarding the methodology of this study and for commenting on the content of this manuscript. The authors are grateful to Gerard Rongen (Radboud university medical center) for reviewing this manuscript.

Contributors HJO originated the idea for this study together with MvL. HJO, JIH and MvL were involved in the design of the study. MvL and JIH

conducted the data analysis and MvL wrote the first draft of the manuscript. JIH and HJO reviewed and revised the manuscript. All authors read and approved the final manuscript.

Funding This research was supported by an unrestricted educational grant from MSD, The Netherlands (Dutch subsidiary of Merck and Co., Inc). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Competing interests HJO is a paid employee from Teva Pharmaceuticals.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Data sets are available on request from the corresponding author.

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

REFERENCES

1. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. 2013. <http://www.icmje.org/icmje-recommendations.pdf>
2. Patel J. Why training and specialization is needed for peer review: a case study of peer review for randomized controlled trials. *BMC Med* 2014;12:128.
3. Black N, van Rooyen S, Godlee F, *et al*. What makes a good reviewer and a good review for a general medical journal? *JAMA* 1998;280:231–3.
4. Schroter S, Black N, Evans S, *et al*. Effects of training on quality of peer review: randomised controlled trial. *BMJ* 2004;328:673.
5. Schroter S, Black N, Evans S, *et al*. What errors do peer reviewers detect, and does training improve their ability to detect them? *J R Soc Med* 2008;101:507–14.
6. Jefferson T, Rudin M, Brodney Folse S, *et al*. Editorial peer review for improving the quality of reports of biomedical studies. *Cochrane Database Syst Rev* 2007;(2):MR000016.
7. Hopewell S, Collins GS, Boutron I, *et al*. Impact of peer review on reports of randomised trials published in open peer review journals: retrospective before and after study. *BMJ* 2014;349:g4145.
8. van Rooyen S, Godlee F, Evans S, *et al*. Effect of blinding and unmasking on the quality of peer review: a randomized trial. *JAMA* 1998;280:234–7.
9. Bordage G. Reasons reviewers reject and accept manuscripts: the strengths and weaknesses in medical education reports. *Acad Med* 2001;76:889–96.
10. Turcotte C, Drolet P, Girard M. Study design, originality and overall consistency influence acceptance or rejection of manuscripts submitted to the Journal. *Can J Anaesth* 2004;51:549–56.
11. Emerson GB, Warne WJ, Wolf FM, *et al*. Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial. *Arch Intern Med* 2010;170:1934–9.
12. Abbot NC, Ernst E. Publication bias: direction of outcome is less important than scientific quality. *Perfusion* 1998;11:182–4.
13. van Lent M, Overbeke J, Out HJ. Role of editorial and peer review processes in publication bias: analysis of drug trials submitted to eight medical journals. *PLoS ONE* 2014;9:e104846.
14. van Lent M, Overbeke J, Out HJ. Recommendations for a uniform assessment of publication bias related to funding source. *BMC Med Res Methodol* 2013;13:120.
15. BMJ. Guidance for peer reviewers. <http://www.bmj.com/about-bmj/resources-reviewers/guidance-peer-reviewers> (accessed March 2014).
16. Groves T. BMJ Peer reviewer training part II: What do editors want from reviewers?. <http://www.bmj.com/sites/default/files/attachments/resources/2011/07/presentation-2what-do-editors-want-reviewers.ppt> (accessed Mar 2014).
17. Diabetologia. Guidance for reviewers. <http://www.diabetologia-journal.org/guidanceforreviewers.html> (accessed Mar 2014).
18. Byrne DW. Common reasons for rejecting manuscripts at medical journals: a survey of editors and peer reviewers. *Sci Ed* 2000;23:39–44.

19. Lippert S, Callaham ML, Lo B. Perceptions of conflict of interest disclosures among peer reviewers. *PLoS ONE* 2011;6:e26900.
20. Mathieu S, Chan AW, Ravaud P. Use of trial register information during the peer review process. *PLoS ONE* 2013;8:e59910.
21. de Vet HC, Mokkink LB, Terwee CB, *et al.* Clinicians are right not to like Cohen's κ . *BMJ* 2013;346:f2125.
22. Bero LA, Rennie D. Influences on the quality of published drug studies. *Int J Technol Assess Health Care* 1996;12:209–37.
23. Angell M. *The truth about the drug companies. How they deceive us and what to do about it.* New York: Random House Trade, 2005.
24. Lundh A, Sismondo S, Lexchin J, *et al.* Industry sponsorship and research outcome. *Cochrane Database Syst Rev* 2012;12:MR000033.
25. Schott G, Pachi H, Limbach U, *et al.* The financing of drug trials by pharmaceutical companies and its consequences. Part 1: a qualitative, systematic review of the literature on possible influences on the findings, protocols, and quality of drug trials. *Dtsch Arztebl Int* 2010;107:279–85.
26. Jones R, Younie S, Macallister A, *et al.* A comparison of the scientific quality of publicly and privately funded randomized controlled drug trials. *J Eval Clin Pract* 2010;16:1322–5.
27. Djulbegovic B, Lacevic M, Cantor A, *et al.* The uncertainty principle and industry-sponsored research. *Lancet* 2000;356:635–8.
28. van Lent M, Rongen GA, Out HJ. Shortcomings of protocols of drug trials in relation to sponsorship as identified by Research Ethics Committees: analysis of comments raised during ethical review. *BMC Med Ethics* 2014;15:83.
29. Mathieu S, Giraudeau B, Soubrier M, *et al.* Misleading abstract conclusions in randomized controlled trials in rheumatology: comparison of the abstract conclusions and the results section. *Joint Bone Spine* 2012;79:262–7.
30. Boutron I, Dutton S, Ravaud P, *et al.* Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 2010;303:2058–64.
31. Boutron I, Altman DG, Hopewell S, *et al.* Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the SPIIN randomized controlled trial. *J Clin Oncol* 2014;32:4120–6.
32. Groves T, on behalf of the BMJ Group. Peer Review—Written evidence submitted by the BMJ Group (PR 41). 2011. <http://www.publications.parliament.uk/pa/cm201011/cmselect/cmsctech/writev/856/m41.htm>
33. Jefferson T, Alderson P, Wager E, *et al.* Effects of editorial peer review: a systematic review. *JAMA* 2002;287:2784–6.
34. Smith R. Peer review: a flawed process at the heart of science and journals. *J R Soc Med* 2006;99:178–82.
35. Godlee F, Gale CR, Martyn CN. Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomized controlled trial. *JAMA* 1998;280:237–40.
36. Kravitz RL, Franks P, Feldman MD, *et al.* Editorial peer reviewers' recommendations at a general medical journal: are they reliable and do editors care? *PLoS ONE* 2010;5:e10072.
37. Mulligan A, Hall L, Raphael E. Peer review in a changing world: an international study measuring the attitudes of researchers. *J Am Soc Info Sci Technol* 2013;64:132–61.

Supplemental Table 1. Checklist for classification of peer review comments

Category	Type of comment
Importance	1. Research question and results of limited clinical relevance/narrow interest
Originality	2. Lack of novelty; paper does not add much to what is already in the published literature
Background and rationale	3. Incorrect/missing background information in introduction 4. Poor justification for conducting the study; insufficient problem statement/rationale provided
Methods	5. Poor experimental design; methodology/study population/data collection methods flawed or questionable 6. Methods inadequately reported; trial design, participants, interventions, outcome measures used, sample size calculation, randomisation, or blinding inaccurately described 7. Statistical analysis methods inappropriate or inadequately reported
Results	8. Results not answering study question, study outcome data incomplete, results inaccurately presented 9. Flow of participants through the study unclear; no explanations provided for non-randomized/excluded subjects or drop-outs
Discussion and conclusion	10. Meaning of study results inadequately discussed; potential explanations and implications insufficiently addressed in discussion 11. Study insufficiently related to prior work in the literature; not discussed how study results support or disagree with previous research 12. Limitations/shortcomings of the study not acknowledged or not sufficiently discussed how they might affect the study results and their interpretation 13. Conclusions inappropriate in relation to study design or results, or overinterpretation of results
Abstract	14. Abstract does not correctly reflect the paper; information in abstract incomplete/inaccurate 15. Discrepancies between data reported in abstract and main text of the paper
References	16. References missing in article, or outdated/irrelevant references used 17. Errors in reference citation; references not appropriate for statements they are meant to support
Presentation	18. Title not representative of the study 19. Poor writing; typos, unclear language, incorrect terminology, text difficult to follow, wordiness, incorrect use of sections in paper 20. Inaccurate/absent/incomplete tables or figures, discrepancies between tables, or discrepancies between main text and tables
Ethics	21. Not clear which research ethics committee provided approval for the study 22. Informed consent procedures inadequately reported, or other ethical issues related to the study
Trial registration, protocol, CONSORT	23. Trial registration absent/incomplete or registration number missing, study protocol not published/provided by authors, or CONSORT statement/diagram missing 24. Deviations from the trial registry or protocol; not adequately explained why these deviations occurred
Conflicts of interest	25. Potential bias introduced by author conflicts of interest, or funder's contribution to the study unclear 26. Systematic bias or spin in the interpretation of results in favour of the study sponsor

Supplemental Table 2. Inter-rater agreement for final classification checklist	
Type of comment	Percentage of agreement
1. Research question of limited clinical relevance	93.1
2. Lack of novelty	94.4
3. Incorrect/missing background information	93.1
4. Poor justification/rationale for conducting study	97.2
5. Poor experimental design	86.1
6. Methods inadequately reported	84.7
7. Statistical analysis methods inappropriate	81.9
8. Study outcome data incomplete	83.3
9. Flow of participants through study unclear	94.4
10. Meaning results inadequately discussed	77.8
11. Study insufficiently related to literature	91.7
12. Limitations not sufficiently discussed	76.4
13. Conclusions inappropriate/overinterpretation	84.7
14. Abstract does not correctly reflect paper	91.7
15. Discrepancies between abstract and main text	98.6
16. References missing/irrelevant references used	95.8
17. Errors in reference citation	95.8
18. Title not representative of study	97.2
19. Poor writing	91.7
20. Inaccurate tables or figures	88.9
21. Ethics committee approval not clear	98.6
22. Other ethical issues related to study	98.6
23. Registration/protocol/CONSORT missing	97.2
24. Deviations from registry or protocol	100.0
25. Bias by author COIs/contribution funder unclear	100.0
26. Systematic bias or spin in favor of sponsor	98.6