Langenbecks Arch Surg (2017) 402:1015–1022
DOI 10.1007/s00423-017-1593-6

SYSTEMATIC REVIEWS AND META-ANALYSES

Journal impact factor and methodological quality of surgical randomized controlled trials: an empirical study

Usama Ahmed Ali [1,2] · Beata M. M. Reiber [1] · Joren R. ten Hove [2] · Pieter C. van der Sluis [2] · Hein G. Gooszen [3] · Marja A. Boermeester [1] · Marc G. Besselink [1]

Received: 22 February 2017 / Accepted: 23 May 2017 / Published online: 4 June 2017

Abstract

Purpose The journal impact factor (IF) is often used as a surrogate marker for methodological quality. The objective of this study is to evaluate the relation between the journal IF and methodological quality of surgical randomized controlled trials (RCTs).

Methods Surgical RCTs published in PubMed in 1999 and 2009 were identified. According to IF, RCTs were divided into groups of low (<2), median (2–3) and high IF (>3), as well as into top-10 vs all other journals. Methodological quality characteristics and factors concerning funding, ethical approval and statistical significance of outcomes were extracted and compared between the IF groups. Additionally, a multivariate regression was performed.

Results The median IF was 2.2 (IQR 2.37). The percentage of 'low-risk of bias' RCTs was 13% for top-10 journals vs 4% for other journals in 1999 ($P < 0.02$), and 30 vs 12% in 2009 ($P < 0.02$). Similar results were observed for high vs low IF groups. The presence of sample-size calculation, adequate generation of allocation and intention-to-treat analysis were independently associated with publication in higher IF journals; as were multicentre trials and multiple authors.

Conclusion Publication of RCTs in high IF journals is associated with moderate improvement in methodological quality compared to RCTs published in lower IF journals. RCTs with adequate sample-size calculation, generation of allocation or intention-to-treat analysis were associated with publication in a high IF journal. On the other hand, reporting a statistically significant outcome and being industry funded were not independently associated with publication in a higher IF journal.

Keywords Impact factor · Methodologic quality and randomized controlled surgical trials

Background

The journal impact factor (IF) is an indicator of how often papers from a certain journal are cited in literature in relation to the number of published papers based on a 3-year period. It was developed by Thompson Reuters and is only reported for journals listed in the Journal Citation Report. The IF's purpose is to measure the influence of a journal in the field, and by extension the influence of articles published within that journal. In daily practice, however, it is often used as a surrogate marker for the quality of both the journal and the individual articles within it.

Previous studies have pointed out the limitations of the IF as an index of methodological quality and have shown that a high IF does not automatically guarantee high quality of individual trials published in such a journal, and vice versa [1, 2]. These analyses only focused on specific journals (most often high IF journal) and lacked broad analyses including all journals within one specific speciality. Moreover, they did not analyse how the reporting of methodology differed between low and high IF journals. Such an analysis could shed more light on the specific

Usama Ahmed Ali and Beata M. M. Reiber contributed equally to this study.

✉ Beata M. M. Reiber
bmmreiber@gmail.com

1 Department of Surgery, G4-190, Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

2 Department of Surgery, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

3 Center for Evidence Based Surgery, Radboud University Nijmegen Medical Center, Geert Grooteplein Zuid 10, 6525 GA Nijmegen, The Netherlands

problems that are encountered by journals of different levels of influence. In addition, other characteristics such as large sample size, multi-center design, significance of the primary outcome and approval of the ethics committee might be important in the association of methodological quality and publication in higher IF journals [3–8]. Therefore, this study aimed to evaluate the association between the journal IF and methodological quality of surgical RCTs as well as to examine the influence of other related characteristics.

Material and methods

Objectives

1. To evaluate the association between the journal IF and methodological quality of surgical RCTs
2. To evaluate the potential association of other factors with publication in a higher IF journal

Search strategy and selection process

We searched PubMed for all surgical RCTs published in two distinct years a decade apart, i.e. 1999 and 2009. First, we searched using the free-text and MeSH term "surgery" with various permutations combined with the Cochrane Highly Sensitive Search strategy. All retrieved hits were subsequently selected according to relevance by two independent reviewers. Inclusion criteria were (1) RCTs, defined as any prospective study assessing the effect of health care interventions in humans randomly allocated to study groups and (2) surgical trials, defined as any trial assessing the effect of a general surgical procedure. Exclusion criteria were (1) non-RCTs and (2) publications in other languages than English, French, German or Dutch. [9].

Assessing reported methodological quality

To evaluate the methodological quality of RCTs, a nine-item list was developed based on the Cochrane guidelines for methodological assessment of randomized trials as published previously [10, 11]. The items were as follows:

1. Stating of explicit primary outcome.
2. Performing a sample-size calculation.
3. Presence of baseline characteristics.
4. Adequate generation of allocation sequence.
5. Concealment of allocation.
6. Presence of blinding (any form).
7. Double-blinding.
8. Intention-to-treat analysis.
9. Adequate handling of dropouts (<20% lost to follow-up, and listing of reasons for all losses).

Additionally, 'low-risk of bias' trials were defined as trials that adequately reported all of the following four items: adequate generation of allocation, adequate concealment of allocation, intention-to-treat analysis and handling of dropouts. The selection of these four criteria was based on empiric evidence indicating the direct influence of these characteristics on effect estimates [12–14].

Study factors

Study, journal and publication characteristics were extracted, including geographical region, number of participants, number of participating centres, impact factor of journal, the type of intervention (surgical, medicines or other), type of funding, reporting of conflict of interest, ethical approval and statistical significance of primary and other outcomes [1, 3–8].

Data and statistical analysis

All results are presented and analysed for both study years (1999 and 2009) separately. Study characteristics and methodological quality criteria are presented and compared between different IF groups. Three IF groups (low, middle and high) were determined by generating tertiles (three equal groups) based on the published RCTs in 2009. For ease of interpretation, the cut-off points were approximated to the closest round number, resulting in the following three groups: low IF (<2), middle (2–3) and high (>3). Additionally, top-10 IF journals were compared to all other journals.

Dichotomous outcomes are presented as the number of events with corresponding percentage, and compared using chi-square test. Continuous data are presented as means with standard deviation (SD) or median and interquartile range (IQR) and were compared by the Student's *t* test or Mann-Whitney *U* test according to normality. For all dichotomous outcomes, risk ratio (RR) with corresponding 95% confidence intervals (95% CI) was presented.

In addition, multivariate linear regression analyses was performed to identify factors independently associated with IF in 2009. For this, only factors showing potential association in univariate analysis were included ($P < 0.2$). Subsequently, we constructed two multivariate models. The first aimed at investigating the association between IF and methodological quality, and included only methodological characteristics as potential predictors. The second additionally included other potentially related factors, as specified under study factors. A backwards regression model was used for both analyses, and a P value of 0.05 was set as a cut-off point for statistical significance. All analyses were performed using IBM SPSS Statistics version 23.

Results

Search results and general characteristics

A total of 12.780 and 25.711 PubMed hits were retrieved on June 3, 2010 for the years 1999 and 2009, respectively [9]. From these, 750 surgical RCTs (300 in 1999 and 450 in 2009) were selected after review by two independent reviewers. Of these, 101 articles were published in journals without an impact factor and were thus classified in the low IF group. The median journal impact factor of published trials was 2.2 (IQR 2.37). Journals ranked in the top-10 in 1999 and 2009 had a median impact factor of 5.39 (IQR 5.91) and 6.32 (IQR 3.82), respectively, compared to 1.87 (IQR 1.41) in 1999 and 2.39 (IQR 1.56) in 2009 for journals not included in the top-10.

The general characteristics for each IF group in both 1999 and 2009 are summarized in Table 1. Trials published in high IF journals originate more often from Europe and North America, are more often multicenter and multinational trials, have a larger sample size and are more often explicitly approved by ethical medical committees. In 1999 and 2009 journals with an IF >3 significantly more often published RCTs that reported industry funding ($P < 0.001$) than other journals. In addition, in 2009, significantly more RCTs published in top-10 journals were industry funded than in other journals (36 vs 15%, respectively, $P < 0.001$).

Methodological quality characteristics

Methodological quality characteristics stratified per IF group and for top-10 journals vs other journals are presented (Tables 2 and 3). For both study years, several important methodological quality characteristics were significantly more often reported in higher IF journal compared to other journals. These included adequate presentation of baseline, performing of a sample-size calculation, adequate stating of the primary aim, generation of allocation sequence and intention-to-treat principle. This resulted in significantly more low-risk of bias RCTs in the high vs low IF group (12 vs 4% in 1999 and 22 vs 8% in 2009), as well as in the top-10 vs other journals (13 vs 4% in 1999 and 30 vs 12% in 2009). Figure 1 provides a graphical depiction of the levels of adequate reporting of the methodological characteristics between the different IF groups in the 2009.

Table 1 Characteristics of included surgical randomized controlled trials

	1999				2009			
	Low IF <2 (*n* = 123)	Med IF 2–3 (*n* = 77)	High IF >3 (*n* = 74)	*P* value	Low IF <2 (*n* = 101)	Med IF 2–3 (*n* = 126)	High IF >3 (*n* = 148)	*P* value
Region								
- Europe	67 (55%)	49 (64%)	38 (51%)	NS	35 (35%)*	62 (49%)	92 (62%)**	<0.0001
- North America	27 (22%)	16 (21%)	21 (28%)	NS	8 (8%)	13 (10%)	27 (18%)	NS
- Australia/Asia	26 (21%)	11 (14%)	15 (20%)	NS	55 (55%)**	41 (33%)*	26 (18%)*	<0.0001
- South America/Africa	3 (2%)	1 (1%)	0 (0%)	NS	3 (2%)	10 (8%)	3 (2%)	NS
Single country trial	113 (92%)	74 (96%)	62 (84%)	NS	95 (94%)**	119(94%)**	120 (81%)*	0.020
Sample size, mean (SD)	117 (18)*	133 (21)*	302 (56)**	<0.001	114 (12)*	103 (12)*	185 (18)**	<0.001
Number of authors, mean (SD)	5 (3)*	6 (3)*	8 (4)**	0.020	6 (3)*	6 (3)*	8 (4)**	0.033
Single centre trial	80 (65%)**	40 (52%)	35 (47%)*	0.002	73 (72%)**	80 (64%)**	66 (47%)*	<0.001
Type of intervention								NS
- Surgical	41 (33%)*	45 (58%)**	21 (28%)*	<0.001	41 (41%)	57 (45%)	75 (51%)	NS
- Medication	52 (42%)	20 (26%)*	44 (60%)**	<0.001	32 (32%)	38 (30%)	50 (34%)	NS
- Other	30 (25%)	12 (16%)	9 (12%)	NS	28 (27%)	31 (25%)	23 (15%)	NS
Published in surgical journal	85 (69%)**	61 (79%)**	28 (38%)*	<0.001	52 (52%)*	89 (71%)**	99 (67%)**	0.007
Primary outcome significant	65 (53%)	43 (56%)	35 (47%)	NS	62 (61%)	79 (63%)	83 (57%)	NS
Any outcome significant	95 (77%)	64 (83%)	63 (85%)	NS	79 (78%)	109 (87%)	116 (79%)	NS
Approval of ethics committee	71 (58%)	48 (62%)	52 (73%)	NS	84 (83%)*	107 (85%)	138 (93%)**	0.029
Type of funding								
- Industry funded	26 (17%)*	16 (21%)	26 (35%)**	<0.001	19 (11%)*	19 (15%)*	45 (30%)**	<0.001
- Other resources/none	28 (19%)	16 (21%)	17 (23%)	NS	63 (36%)	46 (37%)	68 (46%)	NS
- Unclear	95 (64%)**	45 (58%)	31 (42%)*	<0.001	94 (53%)**	61 (48%)**	35 (24%)*	<0.001

Values annotated with ** are significantly higher than values annotated with * ($P < 0.05$)

NS: not significant ($P > 0.05$)

Table 2 Methodological quality characteristics stratified for low, medium and high impact factor

	1999				2009			
	Low IF <2 (n = 149)	Med IF 2–3 (n = 77)	High IF >3 (n = 74)	P value	Low IF <2 (n = 176)	Med IF 2–3 (n = 126)	High IF >3 (n = 148)	P value
1. Baseline presented	129 (87%)*	70 (91%)	73 (99%)**	0.014	156 (89%)*	115 (91%)	143 (97%)**	0.029
2. Sample-size calculation performed	33 (22%)*	29 (38%)**	39 (53%)**	<0.001	55 (31%)*	66 (52%)**	97 (66%)**	<0.001
3. Primary outcome stated	84 (56%)*	61(79%)**	58 (78%)**	<0.001	90 (51%)*	84 (67%)**	123 (83%)***	<0.001
4. Generation of allocation sequence: reported and adequate	39 (26%)*	22 (29%)	35 (47%)**	0.005	67 (38%)*	69 (55%)**	77 (52%)**	0.006
5. Concealment of treatment allocation: reported and adequate	43 (29%)	23 (30%)	30 (41%)	0.190	74 (42%)*	66 (52%)	84 (57%)**	0.024
6. Blinding								
- Performed	51 (34%)	25 (33%)	27 (37%)	0.013	47 (27%)	41 (33%)	50 (34%)	0.653
- Not performed but feasible	31 (21%)	4 (5%)	8 (11%)		28 (16%)	21 (17%)	22 (15%)	
- Not performed, technically difficult	67 (45%)	48 (62%)	39 (53%)		101 (57%)	64 (51%)	76 (51%)	
7. Type of blinding								
- Double blind	33 (22%)	15 (20%)	20 (27%)	0.241	26 (15%)	22 (18%)	28 (19%)	0.595
- Blind assessor/observer	16 (11%)	7 (9%)	6 (8%)		21 (12%)	17 (14%)	26 (18%)	
- Committee for assessment	0 (0%)	0 (0%)	2 (3%)		1 (1%)	1 (1%)	2 (1%)	
- Not stated	100 (67%)	55 (71%)	46 (62%)		128 (73%)	86 (68%)	92 (62%)	
8. Type of analysis								
- Intention-to-treat (adequate)	17 (11%)*	12 (16%)*	31 (42%)**	<0.001	34 (19%)*	42 (33%)**	83 (49%)***	<0.001
- Per protocol analysis	2 (2%)	5 (7%)	2 (3%)		4 (2%)	3 (2%)	9 (6%)	
- Not stated	130 (87%)	60 (78%)	41 (55%)		138 (78%)	81 (64%)	66 (45%)	
9. Handling of dropouts: reported and adequate	111 (75%)*	70 (91%)**	63 (85%)	0.007	136 (77%)	112 (89%)	125 (85%)	0.025
10. Low-risk of bias	6 (4%)*	2 (3%)*	9 (12%)**	0.019	14 (8%)*	20 (16%)	32 (22%)**	0.002

Values annotated with ** are significantly higher than values annotated with * ($P < 0.05$). Values annotated with *** are significantly higher than both other compared groups (either * or **)

Table 3 Methodological quality characteristics stratified for top 10 ranked journals and lower ranked journals

	1999			2009		
	Top 10 ($n = 56$)	Other ($n = 244$)	P value	Top 10 ($n = 69$)	Other ($n = 381$)	P value
1. Baseline presented	54 (96%)	218 (89%)	0.1	68 (99%)*	346 (91%)	0.029
2. Sample-size calculation performed	29 (52%)*	72 (30%)	0.001	51 (74%)*	167 (44%)	<0.001
3. Primary outcome stated	43 (77%)	160 (66%)	0.106	61 (88%)*	236 (62%)	<0.001
4. Generation of allocation sequence: reported and adequate	27 (48%)*	69 (28%)	0.004	42 (61%)*	171 (45%)	0.014
5. Concealment of treatment allocation: reported and adequate	22 (39%)	74 (30%)	0.194	45 (65%)*	179 (47%)	0.005
6. Blinding						
- Performed	22 (39%)	81 (33%)	0.676	22 (32%)	116 (30%)	0.794
- Not performed but feasible	7 (13%)	36 (15%)		9 (13%)	62 (16%)	
- Not performed, technically difficult	27 (48%)	127 (52%)		38 (55%)	203 (53%)	
7. Type of blinding						
- Double blind	15 (27%)	53 (22%)	0.384	12 (17%)	64 (17%)	0.864
- Blind assessor/observer	7 (13%)	22 (9%)		10 (15%)	54 (14%)	
- Committee for assessment	1 (2%)	1 (1%)		0 (0%)	4 (1%)	
- Not stated	33 (59%)	199 (82%)		47 (68%)	259 (68%)	
8. Type of analysis						
- Intention-to-treat (adequate)	24 (43%)*	36 (15%)	<0.001	45 (65%)*	104 (27%)	<0.001
- Per protocol analysis	0 (0%)	9 (4%)		3 (4%)	13 (3%)	
- Not stated	32 (57%)	199 (82%)		21 (30%)	264 (69%)	
9. Handling of dropouts: adequate and reported	49 (88%)	195 (80%)	0.189	59 (86%)	314 (82%)	0.530
10. Low-risk of bias	7 (13%)*	10 (4%)	0.014	21 (30%)*	45 (12%)	<0.001

Values annotated with ** are significantly higher than values annotated with * ($P < 0.05$)

Multivariate analysis

Univariate analysis identified the following factors as potentially correlated to IF: stating of explicit primary outcome, performance of a sample-size calculation, presence of baseline characteristics, adequate generation of allocation sequence, concealment of allocation, intention-to-treat analysis and adequate handling of dropouts (<20% lost to follow-up, and listing of reasons for all losses).

All factors related to methodological quality were included in a multivariate model. This identified the following factors to be independently associated with a higher journal IF:

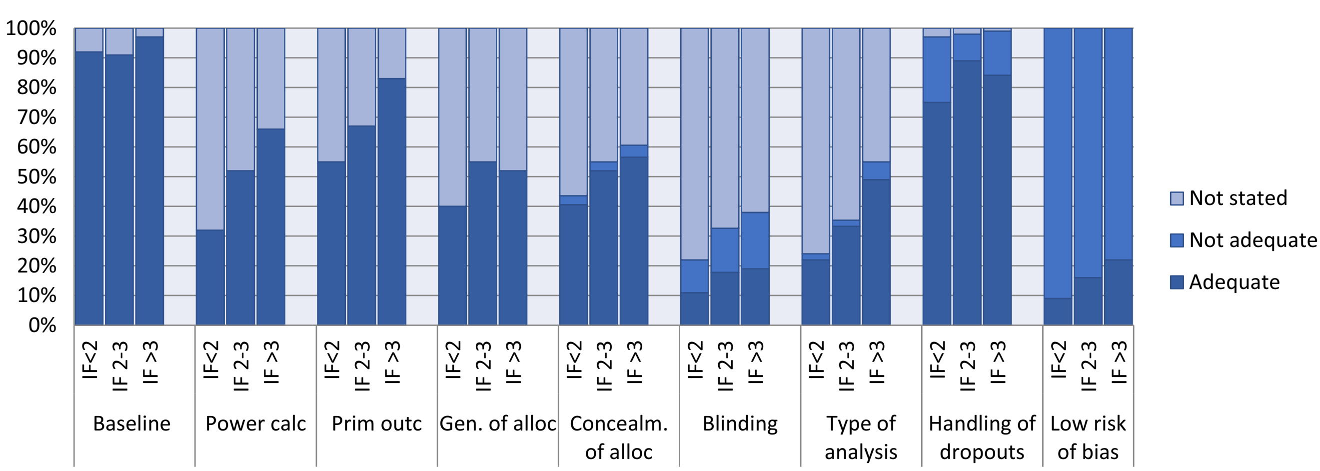

Fig. 1 IF: journal impact factor. Baseline: baseline characteristics presented. Power calc: sample-size calculation performed. Prim outc: primary outcome explicitly stated. Gen of alloc: adequacy of generation of allocation. Concealm of alloc: concealment of allocation. Blinding: presence of blinding. Type of analysis: intention-to-treat analysis performed. Handling of dropouts: adequate handling of dropouts (<20% lost to follow-up, and listing of reasons for all losses). Low-risk of bias: identification as 'low-risk of bias' trial

intention-to-treat analysis, generation of allocation sequence and performance of a sample-size calculation (Table 4).

Subsequently, all potentially related factors were added to the multivariate analysis. This confirmed that intention-to-treat analysis, generation of allocation sequence and performance of a sample-size calculation were independent factors associated with publication in a higher journal IF. In addition, it identified a higher number of authors and multicenter trials as independent factors for higher IF (Table 4). RCTs reporting a statistically significant primary outcome showed a trend towards publication into a higher IF, although this association was not statistically significant.

Discussion

This study identified several important findings. First, we found that RCTs published in a higher IF journal were indeed associated with improved methodological quality as compared to RCTs published in lower IF journals. This was observed both in 1999 and 2009 indicating a consistent finding over time. Secondly, performing an RCT with an adequate sample-size calculation, generation of allocation or intention-to-treat analysis seems to increase the chances of publication in a high IF journal. Third, performing a study within a large cooperative coalition was also significantly associated with publication in a higher IF journal, while finding a statistically significant outcome was not. Finally, despite the observed improvement in methodological quality, the majority of trials still had methodological limitations making individual critical appraisal necessary for all RCTs.

Publication of a surgical RCT in a high IF journal (>3) was associated with a threefold increase in the low-risk of bias status, from 4 to 12% in 1999 and from 8 to 22% in 2009. Similar results were seen in the top-10 journals vs other journal. Still, even in this group, only 30% of published RCTs met the low-risk of bias criteria. Hence, the journal IF can be used as a rough quality indicator but critical appraisal remains necessary for all RCTs [2, 11, 15]. Conversely, we were able to identify three methodological quality criteria that were independently associated with publication in a higher IF journal. All three criteria (i.e. sample-size calculation, adequate generation of randomization sequence and adequate intention-to-treat analyse) are of major importance to the reliability and accuracy of RCTs. Moreover, these criteria are easy to achieve and feasible in virtually all types of (surgical) RCTs. Thus, we hope these results provide an additional stimulus to researchers to take these criteria into consideration when designing and reporting RCTs.

Table 4 Results of regression analysis; independent factors for the journal impact factor

	B-coefficient	95% CI	P
Method characteristics			
Sample-size calculation	1.2	0.4–1.9	0.002
Generation of allocation	1.0	0.3–1.7	0.003
Type of analysis	2.0	1.3–2.8	0.000
All characteristics			
Sample-size calculation	1.0	0.3–1.7	0.007
Generation of allocation	0.8	0.1–1.4	0.024
Type of analysis	1.5	0.7–2.2	<0.001
Multicenter trial	0.7	0.2–1.3	<0.001
Number of authors	0.4	0.3–0.5	<0.001
Significance primary outc	0.6	−0.39–1.3	0.07

Surgical RCTs face a unique set of challenges as compared to pharmaceutical RCTs, as blinding of surgeons and patients is often challenging. Up to 34% of the published surgical trials in our study were blinded in 2009. In contrast, blinding was reported up to 71% in pharmaceutical trials [16]. Even though blinding remains an important measure to prevent performance bias, Probst et al. fairly pointed out that there is no evidence that ornate blinding measures in surgical RCTs yield better evidence [17]. Moreover, surgical placebos entail a considerable risk for study participants and therefore should be used only if justified by the clinical question and methodological necessity [18, 19].

In the multivariate analysis, a trend towards publication of studies with a statistical significant outcome into higher IF journals was observed. Although the association did not reach statistical significance, the existence of such correlation has been extensively discussed in literature with previous studies showing contradicting findings. Gluud et al. reported similar rates of trials with positive primary outcomes (around 70%) in four subgroups of journals stratified by impact factor ($P = 0.32$). On the other hand, Littner et al. found that studies with significant primary outcomes were associated with a slightly higher mean IF than those with negative results (mean IF 4.6 and 4.2, respectively, with a $P = 0.03$). While statistically significant, the scientific relevance of such a small difference is uncertain.

In essence, two main arguments exist for this potential association. Either a publication bias towards positive results driven by fierce competition for publication in high IF journals, or the notion that studies with positive results might be truly better than those with statistically insignificant primary outcomes because of better selection of hypotheses, study design, funding and larger sample-size recruitment [6]. Our results support the notion that such factors related to the quality of trials are indeed more important than the mere reporting of a statistically significant outcome. Similarly, industry funding was not found to be independently associated with higher IF. This topic has been extensively studied by a recent review that found that industry funding of surgical trials leads to exaggerated positive reporting of outcomes [20]. This study also found that industry-funded trials were not published in journals with a higher IF compared with non-industry-funded trials (mean impact factor 4.7 vs 3.2; $P = 0.97$).

Extensiveness of the trial such as large sample size, large number of collaborating authors and multicenter trials were all correlated with publication in high IF journals. Parts of these findings are collaborated by previous studies as well [2, 5]. Several explanations can be formulated. First, for a study to obtain such wide collaboration, it must have a clinical relevant question that makes it interesting for clinical practice. Second, expertise in design and reporting of RCTs are probably more available within a wide collaboration, increasing the quality of the trial. Thirdly, a less illustrious explanation might be that, as Lokker et al. pointed out, when the bigger network is involved, more (self)citations can be obtained [6]. Nonetheless, this finding is important in that it provided clear evidence that putting the time and effort to build the necessary cooperation to perform a well-designed and clinically relevant trial does pay-off in terms of higher appreciation by journals.

Several reports have studied the relation between adequate reporting of conflict of interest and the IF. In our study, despite possessing data on this item, we decided not to include it in the analysis since higher IF journals often have strict and comprehensive procedures for declaring conflict of interest. Therefore, we felt that an association would be inevitable. A recent study evaluating the procedure of disclosing conflict of interest in 64 journals has indeed clearly shown that journals that mandated full disclosure of conflict of interest had a greater IF (0.626 vs 1.732; $P = 0.006$) [21]. In a post hoc analysis, our data do indeed confirm this finding.

Certain choices have been in this study that merit further discussion. Such choice is the inclusion of journals without an IF into the low IF group. This choice can be justified by the fact that most these journals have not been chosen by Thompson Reuters to be included in their Journal Citation Report since they are smaller and less-known journals. However, some of these journals, especially recently launched journals, might not fall into this classification and may develop over the years to become important journals in their respective field. Thus, inclusion of all these journals into the low IF group, while reasonable, might bias the result. Such bias, however, is likely to work against finding differences between low and high IF journals, since 'future' high IF journals are now included in the low IF group. Thus, the effect on the currently observed differences is probably limited. Another choice was to approximate the cut-off points for low, medium and high to the nearest round whole number. While the division is based on creating three equal parts (i.e. tertiles), which is a well-accepted approach, the cut-off points obtained were arbitrary and not very intuitive. To ease readability and interpretation, we have chosen to approximate these cut-off points to be more easily grasped by the reader. This in term resulted in slight inequality of the three groups, which might have affected the results. It is important to note that the choice for approximation was done before any analysis was performed, and that no other grouping was used for analysis.

Strengths of this study include the high number of included trials and the identification of the majority of surgical RCTs in PubMed in two complete years. This offers a comprehensive analysis across the range of all journals. Previous publications have described the possible association between the journal IF and methodological quality, but within considerably smaller numbers of publications, different fields of specialty and less detailed (methodological quality) characteristics retrieved from the publications [2, 4, 15]. Another strength is the inclusion of 2 years, one before and one after the introduction of the CONSORT statement (2001). Thus, the effect observed in this paper also reflects the contribution that the CONSORT statement may have had on the quality of reporting which is consistent with previous publications [22, 23].

This study has some limitations. First, it should be noted that poor reporting does not necessarily indicate poor quality [24]. Nonetheless, the importance of adequate reporting cannot be understated since it is the only tool for clinicians to assess the quality of trials. Secondly, certain factors in the editorial decision of higher IF journals to accept certain RCTs over other, such as novelty of idea and clinical relevance of the research question, are hard to quantify and could thus not be included in this review. However, the importance of these factors is self-eminent and would probably not importantly influence our results. A third limitation is that we were not able to assess 'selective reporting of outcomes', which is part of The Cochrane Risk of Bias tools [25]. To be able to assess this item, reported outcomes of RCTs have to be compared to either their published protocols or to other trails with similar research questions. Both approaches were practically not feasible due to the extensiveness and varying topics of RCTs in this study. Lastly, the search for this review was performed over 6 years ago. Nonetheless, the extensiveness of the search and the time span it covers (a decade apart with the CONSORT statement in between), provides valuable insights that have not been available before, with an effort that is hard to match.

Conclusion

In summary, publication of RCTs in high IF journals is associated with a moderate improvement in methodological quality. Nonetheless, critical appraisal of individual RCTs remains necessary. Performing an RCT with adequate sample-size calculation, generation of allocation or intention-to-treat analysis seems to increase the chances of publication in a high IF journal, as does performing a study within a large cooperative coalition. On the other hand, reporting a statistically significant outcome and being industry funded was not independently associated with publication in a higher IF journal.

Authors' contributions Study conception and design: Usama Ahmed Ali, Beata M. M. Reiber, Hein G. Gooszen, Marja A. Boermeester, Marc G. Besselink; acquisition of data: Beata M.M. Reiber, Usama Ahmed Ali, Joren R. ten Hove, Pieter C. van der Sluis; analysis and interpretation of data: Usama Ahmed Ali, Beata M.M. Reiber, Marja A. Boermeester, Marc G. Besselink; drafting of manuscript: Usama Ahmed Ali, Beata M.M. Reiber; critical revision of manuscript; Usama Ahmed Ali, Beata M.M. Reiber, Joren R. ten Hove, Pieter C. van der Sluis, Hein G. Gooszen, Marja A. Boermeester, Marc G. Besselink.

Compliance with ethical standards

Funding No funding was obtained for this work.

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Not applicable.

References

1. Kurumis AP (2003) Understanding the limitations of the journal impact factor. J Bone Joint Surg Am 85-A(12):2449–2454
2. Gluud L, Sørensen TI, Gøtzsche PC, Gluuc C (2005) The journal impact factor as a predictor of trial quality and outcomes: cohort study of hepatobiliary randomized clinical trials. Am J Gastroenterol 100:2431–2435
3. Olson CM, Rennie D, Cook D, Dickersin K, Flanagin A, Hogan JW, Zhu Q, Reiling J, Pace B (2002) Publication bias in editorial decision making. JAMA 287:2825–2828
4. Bridoux V, Moutel G, Roman H, Kianifard B, Michot F, Herve C, Tuech JJ (2012) Methodological and ethical quality of randomized controlled clinical trials in gastrointestinal surgery. J Gastrointest Surg 16:1758–1767
5. Kulkarni A, Busse JW, Shams I (2007) Characteristics associated with citation rate of the medical literature. PLoS One 2(5):e403
6. Lokker C, Haynes RB, Chu R, McKibbon KA, Wilczynski NL, Walter SD (2012) How well are journal and clinical article characteristics associated with the journal impact factor? A retrospective cohort study. J Med Libr Assoc 100(1):28–33
7. Littner Y, Mimouni FB, Dollberg S, Mandl D (2005) Negative results and impact factor. Arch Pediatr Adolesc Med 159:1036–1037
8. Okike K, Kocher MS, Torpey J, Nwachukwu BU, Mehlman CT, Bhandari M (2011) Level of evidence and conflict of interest disclosure associated with higher citation rates in orthopedics. J Clin Epidemiol 64:331e338
9. Ahmed Ali U, van der Sluis PC, Issa Y, Abou Habaga I, Gooszen HG, Flum DR, Algra A, Besselink MG (2013) Trends in Worldwide Volume and Methodological Quality of Surgical Randomized Controlled Trials. Ann Surg 258:199–207
10. Higgins J, Green S (2011) Cochrane handbook for systematic reviews of interventions version 5.1.0. The Cochrane Collaboration
11. Chan AW, Altman DG (2005) Epidemiology and reporting of randomized trials published in PubMed journals. Lancet 365:1159–1162
12. Juni P, Altman DG, Egger M (2001) Systematic reviews in health care: assessing the quality of controlled clinical trials. BMJ 323:42–46
13. Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, Gluud C, Martin RM, Wood AJG, Sterne JAC (2008) Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 336:601–605
14. Schulz KF, Chalmers I, Hayes RJ, Altman DG (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 273:408–412
15. Sinha S, Sinha A, Ashby E, Jayaram R, Grocott MP (2009) Quality of reporting in randomized trials published in high quality surgical journals. J Am Coll Surg 209:565–571
16. Boutron I, Estellat C, Guittet L, Dechartres A, Sackett D, Hróbjartsson A, Ravaud P (2006) Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. PLoS Med 3(10):e425
17. Probst P, Grummich K, Heger P, Zaschke S, Knebel P, Ulrich A, Büchler MW, Diener MK (2016) Blinding in randomized controlled trials in general and abdominal surgery: protocol for a systematic review and empirical study. Syst Rev 24(5):48
18. Boutron I, Tubach F, Giraudeau B, Ravaud P (2004) Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. J Clin Epidemiol 57(6):543–550
19. Probst P, Grummich K, Harnoss JC, Hüttner FJ, Jensen K, Braun S, Kieser M, Ulrich A, Büchler MW, Diener MK (2016) Placebo-controlled trials in surgery. Medicine (Baltimore) 95(17):e3516
20. Probst P, Knebel P, Grummich K, Tenckhoff S, Ulrich A, Büchler MW, Diener MK (2016) Industry bias in randomized controlled trials in general and abdominal surgery: an empirical study. Ann Surg 264(1):87–92
21. Probst P, Hüttner FJ, Klaiber U, Diener MK, Büchler MW, Knebel P (2015) Thirty years of disclosure of conflict of interest in surgery journals. Surgery 157(4):627–633
22. Kane RL, Wang J, Garrard J (2007) Reporting in randomized clinical trials improved after adoption of the CONSORT statement. J Clin Epidemiol 60:241–249
23. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, Gaboury I (2006) Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Med J Aust 185:263–267
24. Devereaux PJ, Choi PT, El-Dika S, Bhandari M, Montori VM, Schünemann HJ, Garg AX, Busse JW, Heels-Ansdell D, Ghali WA, Manns BJ, Guyatt GH (2004) 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 57:1232–1236
25. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG (2004) Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 291(20):2457–2465