Rheumatol Int
DOI 10.1007/s00296-014-2963-9

ORIGINAL ARTICLE

Analysis of randomized controlled trials in *Rheumatology International* from 1981 to 2012: methodological assessment

Jeong Woo Lee · Jae Hoon Chung · Jung Ki Jo · Seung Wook Lee

Received: 16 November 2013 / Accepted: 4 February 2014

Abstract The aim of the study is to assess the methodological quality of randomized controlled trials (RCTs) published in *Rheumatology International* (*RI*) by using three types of analytical tools. MEDLINE was used to extract RCTs from original articles published in the *RI* from 1981 (vol. 1) to 2012 (vol. 32). The relationship between the number of articles and RCTs with time and that between various factors and the quality of RCTs were analyzed. To analyze the methodological quality of the RCTs, the time period was divided into several sections and three tools were applied (e.g., the Jadad scale, van Tulder scale, and Cochrane Collaboration Risk of Bias Tool). The number of RCTs published gradually increased with time significantly ($p < 0.001$). The differences in RCT quality scores by each method in the publication years evaluated were not statistically significant, but RCTs that included descriptions of allocation concealment methods had received institutional review board (IRB) approval, and that conducted in the multicenter had significantly higher-quality scores than other studies. In conclusion, although the number of RCTs published in *RI* since its publishing in 1981 has increased with time, but no qualitative improvement of RCT was observed over time. It is necessary to improve the reporting of concealment of allocation, generation of randomization sequences, design of blinded studies, and obtaining IRB approval, all of which are criteria of high-quality RCTs.

J. W. Lee · S. W. Lee (✉)
Department of Urology, Guri Hospital, Hanyang University, 249-1, Gyomun-dong, Guri 471-701, Korea
e-mail: swleepark@hanyang.ac.kr

J. W. Lee · J. H. Chung · J. K. Jo · S. W. Lee
Department of Urology, College of Medicine, Hanyang University, Seoul, Korea

Keywords Randomized, controlled trial · Research design · Data quality · Journal article

Introduction

Since the introduction of evidence-based medicine (EBM), there has been a gradual recognition of the importance of the clinical application of objective outcomes derived from well-planned research. The randomized, clinical trials (RCTs) are considered as a research method that can obtain reliable data with reduced experimental bias and a high level of evidence [1]. The reporting of RCTs must achieve a high standard to enable physicians to measure the validity of the results [2], and the methodology of RCTs must be sound to avoid incorrect conclusions because of study biases [3].

To minimize such errors, the International Committee of Medical Journal Editors (ICMJE) recommends following the Consolidated Standards of Reporting Trials (CONSORT) guidelines when reporting RCT results [4]. These guidelines enable researchers to use appropriate research methods and to conduct studies and analyze the results appropriately [5]. Also, the CONSORT guidelines help physicians understand and evaluate research methods, data analysis, and trial results. Consequently, the guidelines reinforce the importance of evaluating methodological quality in the design, conduct, and analysis of RCTs.

However, the CONSORT checklist does not allow for quantifying each item, which limits its use for comparison of RCTs. Quality assessment methods based on numerical scale scoring systems are useful for evaluating the quality of clinical trials. They also enable easy comparison of trial quality. The Jadad, van Tulder, Cochrane Collaboration Risk of Bias Tool (CCRBT), Newell's,

Published online: 25 February 2014

Scottish Intercollegiate Guidelines Network (SIGN), and the National Institute for Health and Clinical Excellence (NICE) scales have all been applied to the methodological assessment of RCTs [6–8].

The Jadad scale has comparatively simple assessment items, which makes it easy to use; however, it does not include an assessment of "allocation concealment." Effective allocation concealment is essential to prevent selection bias in the allocation of patients to study treatment arms. The van Tulder scale and CCRBT include assessment items for allocation concealment.

The above tools have been used by Chung and Lee [9] and Cho et al. [10], to evaluate the quality of RCTs published in the *Korean Journal of Urology* and the *International Journal of Urology* (*IJU*). The *Rheumatology International* (*RI*) has a long, distinguished history focusing on rheumatology, and this study assessed the methodological quality of RCTs published in *RI* using the Jadad scale, van Tulder scale, and CCRBT, to further the future research paradigm of "Rheumatology."

Materials and methods

Articles and extraction methods for RCTs

A total of 4,194 articles published in *RI*, from 1981 (vol. 1) to 2012 (vol. 32), were retrieved using a manual Web search of the MEDLINE journal database. Publication type was filtered by "randomized, controlled trial." Additional searches were conducted using keywords including "randomized," "randomization," and "randomly," to identify any omitted articles. RCTs were chosen from articles reported in Index Medicus vol. 1, 1981. To ensure the objectivity of the study, two reviewers who had been trained in RCT methodology independently performed data extraction. The extraction results of the two reviewers were coordinated and incorporated into one dataset. Any disagreements were discussed by the three reviewers and a consensus was reached.

Quality validation tools and assessment methods

Analyses were performed independently by two reviewers using the Jadad scale, van Tulder scale, and CCRBT. The results were coordinated by a third reviewer.

Jadad scale

The Jadad scale (also known as the Oxford quality scoring system) has a maximum score of five points. Two points are awarded for the description of the randomization method, two points for the blinding method, and one point for including the information of study dropouts. If mention of randomization and blinding was included without any description of the methods used, one point was assigned for each. When a description of an adequate method was included, one point was added. Two points for each factor mean appropriate randomization and blinding. One point was added if the number of dropouts in each study group and the reasons were specified. If there were no dropouts, then this should have been stated. RCT reports with a Jadad score of three points or more were selected as high-quality studies. Studies in which double-blinding was not possible, but had a total score of two points or more, were also selected as high quality [11].

vanTulder scale

The van Tulder scale assessed 11 components, including randomization, concealment of allocation, baseline characteristics, three blinding methods, co-intervention, compliance, dropouts, endpoint assessment time point, and intention-to-treat analysis. The van Tulder assessment method involved the selection of "yes," "no," or "don't know" for each item. When at least five criteria were satisfied (five points or more), the RCT was assessed as high quality [12].

CCRBT

The CCRBT assessed six domains: random sequence generation, allocation concealment, blinding, incomplete outcome data, selective outcome reporting, and other potential threats to validity. For each domain, the CCRBT assessment involved the selection of "yes," "no," or "unclear." After the six domains were analyzed, if the first three domains were "yes," and no important concerns related to the last three domains were identified, then the study was classified as having a low risk of bias. Studies assessed as having two or fewer domains with "unclear" or "no" responses were classified as having a moderate risk of bias. Studies assessed as having three or more domains with "unclear" or "no" responses were classified as having a high risk of bias [13].

Analysis of RCT characteristics

The effects of allocation concealment, institutional review board (IRB) approval, and funding support on RCT quality were analyzed, and the impact of classification as a single-center or multicenter study was assessed. Classification by country was analyzed based on the affiliation of the corresponding author.

Statistical analysis

One-way ANOVA was used to compare differences of scores obtained with each assessment tool. Comparisons of

the numbers and percentages of high-quality RCT articles are published in each time period, and quality assessments with the CCRBT scale were performed with the Chi-square test. Differences in the score according to RCT characteristics were analyzed using Student's *t* test. SPSS version 18.0 was used for all statistical analyses, and a *p* value of <0.05 was considered statistically significant.

Results

Inter-rater reliability scores between reviewers

A total of 4,194 articles published in *RI*, from 1981 to 2012, were retrieved by searching the MEDLINE database. Of these, 128 articles (3.05 %) were RCTs and were subjected to subsequent analyses. The rate of concordance between the Jadad scale and van Tulder scale was very high: 97 and 92 %, respectively. CCRBT, a more complex assessment method, showed an 84 % concordance rate.

Quantitative variation in RCTs over time

The time period was divided into six groups for analysis: years 1981–1987, 1988–1992, 1993–1997, 1998–2002, 2003–2007, and 2008–2012. In each time period, the number of RCTs (% of papers that were RCTs) was 3 (1.03 %), 7 (3.06 %), 12 (5.63 %), 15 (5.14 %), 43 (4.92 %), and 48 articles (2.22 %), respectively ($p < 0.001$) (Table 1). The number of RCTs published increased with time significantly.

Quality assessment of RCT articles over time

Jadad scale

The Jadad scores for each time period were 2.00 ± 0.00, 3.29 ± 0.95, 2.50 ± 0.52, 2.33 ± 1.18, 2.72 ± 1.12, and 2.29 ± 1.20 points, respectively ($p = 0.168$). The numbers (%) of high-quality RCTs in each time period were 0 (0 %), 6 (85.71 %), 6 (50 %), 7 (46.67 %), 24 (55.81 %), and 18 articles (37.50 %), respectively ($p = 0.138$) (Table 2).

Table 1 Characteristics of RCTs according to publication year

Year	Original articles	RCT (%)	Concealment of allocation (%)	Appropriate randomization (%)	IRB approval (%)	Blind study (%)	Appropriate blindness (%)	Drug study (%)
1981–1987	291	3 (1.03)	0	0	0	2 (66.67)	0	1 (33.33)
1988–1992	229	7 (3.06)	2 (28.57)	2 (28.57)	4 (57.14)	7 (100)	2 (28.57)	6 (85.71)
1993–1997	213	12 (5.63)	4 (33.33)	1 (8.33)	5 (41.67)	7 (58.33)	2 (16.67)	9 (75)
1998–2002	292	15 (5.14)	4 (26.67)	4 (26.67)	2 (13.33)	9 (60)	2 (13.33)	6 (40)
2003–2007	874	43 (4.92)	9 (20.93)	10 (23.26)	23 (53.49)	31 (72.09)	11 (25.58)	25 (58.14)
2008–2012	2167	48 (2.22)	7 (14.58)	16 (33.33)	35 (72.92)	23 (47.92)	6 (12.50)	13 (27.08)
p value		<0.001	0.597	0.472	<0.001	0.068	0.537	0.003
Total	4,194	128 (3.05)	26 (20.31)	33 (25.78)	69 (53.90)	79 (61.72)	23 (17.97)	60 (46.88)

p value was calculated using the Chi-square test

Table 2 Quality assessment of RCTs by publication year

Publication year	Jadad scale		Van Tulder scale		Cochrane's assessment of risk bias		
	Score	High quality	Score	High quality	Low risk	Moderate risk	High risk
1981–1987	2.00 ± 0.00	0 (0 %)	3.67 ± 0.58	0 (0 %)	0	0	3 (100 %)
1988–1992	3.29 ± 0.95	6 (85.71 %)	5.43 ± 2.07	5 (71.43 %)	1 (14.29 %)	1 (14.29 %)	5 (71.42 %)
1993–1997	2.50 ± 0.52	6 (50 %)	5.17 ± 1.53	7 (58.33 %)	0	3 (25 %)	9 (75 %)
1998–2002	2.33 ± 1.18	7 (46.67 %)	4.53 ± 1.73	7 (46.67 %)	1 (6.67 %)	3 (20 %)	11 (73.33 %)
2003–2007	2.72 ± 1.12	19 (44.19 %)	5.16 ± 2.06	24 (55.81 %)	3 (6.98 %)	7 (16.28 %)	33 (76.74 %)
2008–2012	2.29 ± 1.20	18 (37.50 %)	4.56 ± 1.99	23 (47.92 %)	4 (8.33 %)	2 (4.17 %)	42 (87.5 %)
p value	0.168	0.138*	0.465	0.388*	–	–	0.615*
Total	2.51 ± 1.12	56 (43.75 %)	4.84 ± 1.93	66 (51.56 %)	9 (7.03 %)	16 (12.5 %)	103 (80.47 %)

p value was calculated using one-way ANOVA test

* Chi-square test

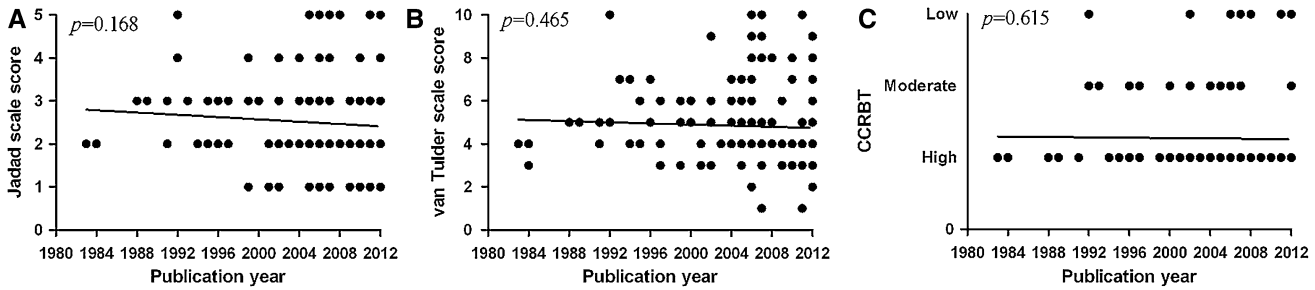

Fig. 1 The result of methodological quality assessment of RCTs with publication year by using the Jadad scale (**a**), van Tulder scale (**b**), and CCRBT (**c**). *CCRBT* Cochrane Collaboration Risk of Bias Tool, *Low* low risk of bias, *Moderate* moderate risk of bias, *High* high risk of bias

No statistically significant differences in the RCT quality scores were observed for the publication years that were evaluated (Fig. 1a). The appropriate blind studies were assessed by 23 researches and accounted for 17.97 % of 128 RCTs and 29.11 % of 79 blind studies.

vanTulder scale

Assessment scores by the van Tulder scale for each time period were 3.67 ± 0.58, 5.43 ± 2.07, 5.17 ± 1.53, 4.53 ± 1.73, 5.16 ± 2.06, and 4.56 ± 1.99 points, respectively ($p = 0.465$), with 0 (0 %), 5 (71.43 %), 7 (58.33 %), 7 (46.67 %), 24 (55.81 %), and 23 articles (47.92 %) being assessed as high quality, respectively ($p = 0.388$) (Table 2). The differences in RCT quality scores in the publication years evaluated were not statistically significant (Fig. 1b).

CCRBT

The CCRBT identified 0 (0 %), 1 (14.29 %), 0 (0 %), 1 (6.67 %), 3 (6.98 %), and 4 articles (8.33 %) in each time period, respectively, as having a low risk of bias ($p = 0.615$) (Table 2). The CCRBT assessments for each publication time interval were not significantly different (Fig. 1c).

Analysis of other factors related to RCT quality

RCT reports that included descriptions of allocation concealment methods (Jadad 3.65 ± 1.06, van Tulder 7.23 ± 1.78; $p < 0.001$, $p < 0.001$, respectively) had received IRB approval (2.71 ± 1.24, 5.17 ± 2.15; $p = 0.023$, 0.032), and that conducted in the multicenters (2.73 ± 1.09, 5.21 ± 2.17; $p = 0.032$, 0.039) had significantly higher-quality scores than other studies. There were no quality score differences between articles with regard to the type of intervention (drug study: 2.57 ± 0.91, 5.15 ± 1.96; $p = 0.569$, 0.092), funding support (2.48 ± 0.85, 4.43 ± 1.70; $p = 0.889$, 0.264) (Table 3). The affiliation of the corresponding author has been diverse every decade since 1981 (Table 4).

Discussion

This methodological quality assessment of RCTs that were published in the *RI* from 1981 to 2012 revealed significant increases in number over this time period, but there was no significant improvement in terms of RCT quality with time. Scales et al. [14] reported that the number of RCTs increased over time when they compared RCTs published in the *Journal of Urology*, *Urology*, *European Urology*, and the *British Journal of Urology International* in 1996 and 2004. Similarly, a study by Cho et al. [10] in the IJU reported that the number of RCTs has increased over the last 18 years. In the present study, only three RCTs published from 1981 to 1987 were examined; however, the number gradually increased to a total of 48 RCTs published from 2008 to 2012. This result is likely a consequence of the increasing importance of EBM in current medical development, and RCTs are considered as representing the highest level of evidence [15].

Most published assessments of RCT quality, using various methods, have verified that a majority of RCTs were of low quality. Bridoux et al. [16] reported suboptimal quality of RCTs in gastrointestinal surgery. Cho et al. [10] analyzed annual publication of RCTs in IJU and found no improvement in quality over the time period that was studied. Likewise, this evaluation of the RCTs published in *RI* revealed no improvement in quality with time. It is supposed that RCT is known the most reliable study in EBM, but awareness about the reporting methods used by researchers is rather poor. Therefore, even though the research may be performed as per the protocol, poor reporting style may lead to a low score for the methodology.

In the Jadad scale, research papers could be assessed only by the elements of randomization, blinding, and dropout reporting [11]. This scale was easy to use because the quality of research papers was assessed by only three criteria, but it was limited by the lack of accounting for the allocation of concealment. Therefore, the van Tulder scale and the CCRBT were used in this study to compensate for this problem. The van Tulder scale is more comprehensive

Table 3 Analysis of factors related to the quality of RCTs

Factors	No. of RCTs (%)	Jadad scale		Van Tulder scale		Cochrane's assessment of risk bias		
		Score	High quality (%)	Score	High quality (%)	High risk (%)	Moderate risk (%)	Low risk (%)
Concealment of allocation								
Described	26 (25.5)	3.65 ± 1.06	22 (84.6)	7.23 ± 1.78	26 (100)	5 (19.2)	12 (46.2)	9 (34.6)
Not described	102 (74.5)	2.22 ± 0.93	34 (33.3)	4.24 ± 1.44	40 (39.2)	98 (96.1)	4 (3.9)	0
p value		<0.001	<0.001*	<0.001	<0.001*			<0.001*
IRB								
Approved	69 (53.9)	2.71 ± 1.24	34 (49.3)	5.17 ± 2.15	40 (58.0)	52 (75.4)	9 (13.0)	8 (11.6)
Not approved	59 (46.1)	2.27 ± 0.91	22 (37.3)	4.46 ± 1.58	26 (44.1)	51 (86.4)	7 (11.9)	1 (1.7)
p value		0.023	0.212*	0.032	0.156*			0.084*
Intervention								
Drug study	60 (46.9)	2.57 ± 0.91	27 (45.0)	5.15 ± 1.96	34 (56.7)	50 (83.3)	7 (11.7)	3 (5.0)
Non-drug study	68 (53.1)	2.46 ± 1.28	29 (42.6)	4.57 ± 1.88	32 (47.1)	53 (77.9)	9 (13.2)	6 (8.9)
p value		0.569	0.859*	0.092	0.293*			0.657*
Funding								
Supported	23 (18.0)	2.48 ± 0.85	10 (43.5)	4.43 ± 1.70	11 (47.8)	18 (78.3)	5 (21.7)	0
Not supported	105 (82.0)	2.51 ± 1.17	46 (43.8)	4.93 ± 1.98	55 (52.4)	85 (81.0)	11 (10.5)	9 (8.5)
p value		0.889	1.000*	0.264	0.819*			0.142*
Center								
Multicenter study	62 (48.4)	2.73 ± 1.09	33 (53.2)	5.21 ± 2.17	36 (58.1)	45 (72.6)	12 (19.4)	5 (8.0)
Single-center study	66 (51.6)	2.30 ± 1.11	23 (34.8)	4.50 ± 1.63	30 (45.5)	58 (87.8)	4 (6.1)	4 (6.1)
p value		0.032	0.050*	0.039	0.162*			0.060*

p value was calculated using Student's *t* test

* Chi-square test

Table 4 Distribution of the nationality of the corresponding authors

	Publication year			Total	*p* value
	1981–1992	1993–2002	2003–2012		
The affiliation of the corresponding author					
Turkey	0	5 (18.5 %)	46 (50.5 %)	51 (39.8 %)	
Germany	3 (30.0 %)	4 (14.8 %)	10 (11.0 %)	17 (13.3 %)	
Italy	0	0	9 (9.9 %)	9 (7.0 %)	
Spain	0	2 (7.4 %)	5 (5.5 %)	7 (5.5 %)	
China	0	2 (7.4 %)	3 (3.3 %)	5 (3.9 %)	
USA	1 (10.0 %)	1 (3.7 %)	2 (2.2 %)	4 (3.12 %)	
Israel	0	3 (11.1 %)	1 (1.1 %)	4 (3.12 %)	
United Kingdom	1 (10.0 %)	0	2 (2.2 %)	3 (2.3 %)	
Finland	0	3 (11.1 %)	0	3 (2.3 %)	
Others	5 (50.0 %)	7 (25.9 %)	13 (14.3 %)	25 (19.5 %)	
Total	10	27	91	128	<0.001
No. of countries	7	14	18	26	

Others: South Korea, Hungary, Australia, Japan, India, Iran, Brazil, the Netherland, Poland, Austria, Ireland, Switzerland, Sweden, Serbia, Denmark, New Zealand, and Greece

because it includes 11 elements; the addition of the CCRBT, which adds an assessment of the risk of study bias, increases the reliability of the results through the use of three complementary assessment tools. Therefore, the results of the evaluation by these methods can differ from each other. However, this study did not show the significant increase in the scores assessed by all of the assessment methods as well as the number of high-quality RCTs.

Scales et al. [14] emphasized randomization and the use of adequate blinding as basic items that should be components of high-quality RCT articles. Well-designed randomization helps to reduce selection bias. A perfect randomization process has two stages: generating an unpredictable random sequence and maintaining concealment of allocation for participants during the study [17, 18]. If the selected study participants recognize the random sequence or determine their allocation, they will change their study behavior, with consequent impact on treatment efficacy and study results. Therefore, these stages are important processes for the appropriate implementation of a clinical trial [19]. And, blinding methods apply to the study participants, health care providers, and those assessing the outcomes [20]. Each blinding method is intended to prevent biases that can be generated during the study.

Mills et al. [21] found that appropriate randomization generation/allocation concealment protocols were reported in 78.4/47.1, 90.5/42.9, 59.1/22.7, 90.1/63, and 74.4/42.3 % of RCTs published in *BMJ*, *Annals of Internal Medicine*, *Archives of Internal Medicine*, *The Lancet*, and *JAMA*, respectively. They were generally insufficient to describe the process of allocation concealment. Additionally, the results of a review by Peron et al. [22] that assessed all the RCTs published in medical oncology journals from 2006 to 2009 showed that only 29 and 51 % of the trials accurately described the processes of randomization and allocation concealment, respectively. In *RI*, there were only 26 RCTs (20.31 %) that properly described the concealment of allocation and 33 RCTs (25.78 %) that described appropriate randomization methods. These rates are lower than those reported for the other journals mentioned above, showing the need for improvement.

Scales et al. reported that the < 50 % of RCTs published in four different urology journals in 2004 had proper descriptions of blinding methods [14]. Chan et al. [19] performed a cross-sectional analysis of 519 articles published in December 2000 and identified in PubMed as RCTs. Among those articles, 309 (60 %) stated that they had used blinding methods, but only 148 of them (48 %) provided a proper description of the blinding procedures. In this study, there were 79 blinded studies (61.72 %), a higher percentage than in other reports. However, only 23 RCT articles (29.11 %) described the blinding design in detail.

As an international standard for RCTs, the IRB review process involves recognition of the validity of the study protocol and is obtained before study implementation. To pass IRB review, a study protocol should be designed appropriate so that it can answer the experimental questions. This helps to improve the study quality and to prove the feasibility of the study [23]. Therefore, RCTs that received IRB approval are expected to be of outstanding quality compared with those that have not received it. Indeed Bridoux et al. [16] reported that most high-quality RCTs have received IRB approval. In this study, the Jadad and van Tulder scale results showed that quality scores of RCTs that had received IRB approval were high and included many research reports assessed as having high quality.

According to Jo et al. [24], which performed the quality assessment in the Journal of endourology, the number of countries increased over time and has raised global awareness. In the present study, we found that the affiliation of the corresponding author was limited to certain countries. Studies from 26 countries were published, among which 53 % were based in Turkey and Germany, but, as the Jo et al., on a 10-year basis, the nationality of corresponding authors increased from 7 (1981–1992) to 14 (1993–2002) and then to 18 (2003–2012) RCTs. The results suggest that this journal has observed as increase in its reputation worldwide.

Peron et al. [22] stated two reasons why articles lack accurate descriptions of methodology. First, because most researchers are clinicians, they focus mainly on the clinical features rather than the methodological items presented in research papers. Second, although there may not be problems with the study design and execution, limitations on the length of journal articles prompt the researcher to include less information. However, there are other studies concluding that poor-quality RCT reports are mainly the result of poor methodology [25] and are likely to reflect study bias and to exaggerate treatment efficacy [26]. Therefore, not only researchers, but also clinicians, should be reminded of the importance of methodological aspects for the highest quality in RCTs.

This study had some limitations. It is possible that the subjective judgment of the investigators influenced the extraction of the RCT characteristics and the quality assessment. To ensure the objectivity and reliability of this study, two reviewers performed the analysis independently. If there were any difference in their assessment results, a third objective reviewer performed a reassessment and adjusted the results. Additionally, although we used well-known quality assessment tools for RCTs, not all items listed in the CONSORT statement were considered. Nevertheless, the use of three different validated tools meant that the majority of items were considered and the conclusions drawn from the analysis were strengthened.

Conclusion

Although the number of RCTs published in *RI* since its publishing in 1981 has increased with time, no qualitative improvement of RCT was observed over time. It is necessary to improve the reporting of concealment of allocation,

generation of randomization sequences, design of blinded studies, and obtaining IRB approval, all of which are criteria of high-quality RCTs. Therefore, for the continued advancement of *RI*, more attention should be directed to appropriate methodological protocols beginning with study implementation. Consequently, our results may suggest a direction that will be useful for improving the quality of future study articles for rheumatologists.

Acknowledgments This work was supported by the research fund of Hanyang University (HY-201100000000250).

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

References

1. Hatala R, Keitz S, Wyer P, Guyatt G (2005) Tips for learners of evidence-based medicine: 4. Assessing heterogeneity of primary studies in systematic reviews and whether to combine their results. Can Med Assoc J 172:661–665
2. Altman DG (1996) Better reporting of randomised controlled trials: the CONSORT statement. BMJ 313:570–571
3. Jackson JL, Srinivasan M, Rea J, Fletcher KE, Kravitz RL (2011) The validity of peer review in a general medicine journal. PLoS One 6:e22475
4. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ et al (2012) CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 10:28–55
5. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ et al (2010) CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 340:c869
6. Newell SA, Sanson-Fisher RW, Savolainen NJ (2002) Systematic review of psychological therapies for cancer patients: overview and recommendations for future research. J Natl Cancer Inst 94:558–584
7. Harbour R, Miller J (2001) A new system for grading recommendations in evidence based guidelines. BMJ 323:334–336
8. Hill J, Bullock I, Alderson P (2011) A summary of the methods that the National Clinical Guideline Centre uses to produce clinical guidelines for the National Institute for Health and Clinical Excellence. Ann Intern Med 154:752–757
9. Chung JH, Lee SW (2013) Assessing the quality of randomized controlled urological trials conducted by Korean medical institutions. Korean J Urol 54:289–296
10. Cho HJ, Chung JH, Jo JK, Kang DH, Cho JM, Yoo TK, Lee SW (2013) Assessments of the quality of randomized controlled trials published in International Journal of Urology from 1994 to 2011. Int J Urol. doi:10.1111/iju.12150
11. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ (1996) Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 17:1–12
12. van Tulder M, Furlan A, Bombardier C, Bouter L (2003) Updated method guidelines for systematic reviews in the Cochrane collaboration back review group. Spine (Phila Pa 1976) 28:1290–1299
13. Higgins JP, Green S (eds) (2011) Cochrane handbook for systematic reviews of interventions. Ver. 5.1.0. The Cochrane collaboration (7 Jan 2011). Available from: http://www.cochrane-handbook.org
14. Scales CD Jr, Norris RD, Keitz SA, Peterson BL, Preminger GM, Vieweg J, Dahm P (2007) A critical assessment of the quality of reporting of randomized, controlled trials in the urology literature. J Urol 177:1090–1094
15. Ioannidis JP (2005) Why most published research findings are false. PLoS Med 2:e124
16. 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
17. 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
18. Altman DG (1991) Randomisation. BMJ 302:1481–1482
19. Chan AW, Altman DG (2005) Epidemiology and reporting of randomised trials published in PubMed journals. Lancet 365:1159–1162
20. Schulz KF, Altman DG, Moher D (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 340:c332
21. Mills EJ, Wu P, Gagnier J, Gagnier J, Devereaux PJ (2005) The quality of randomized trial reporting in leading medical journals since the revised CONSORT statement. Contemp Clin Trials 26:480–487
22. Peron J, Pond GR, Gan HK, Chen EX, Almufti R, Maillet D, You B (2012) Quality of reporting of modern randomized controlled trials in medical oncology: a systematic review. J Natl Cancer Inst 104:982–989
23. Chung JH, Lee JW, Jo JK, Kim KS, Lee SW (2013) A quality analysis of randomized controlled trials about erectile dysfunction. World J Mens Health 31:157–162
24. Jo JK, Autorino R, Chung JH, Kim KS, Lee JW, Baek EJ, Lee SW (2013) Randomized controlled trials in endourology: a quality assessment. J Endourol 27:1055–1060
25. Huwiler-Muntener K, Juni P, Junker C, Egger M (2002) Quality of reporting of randomized trials as a measure of methodologic quality. JAMA 287:2801–2804
26. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG et al (2008) Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 336:601–605