ORIGINAL RESEARCH—EDUCATION

Reporting of Randomized Controlled Trials in Andrology Journals: A Quality Assessment

Jung Ki Jo, MD,* Jae Hoon Chung, MD,† Kyu Shik Kim, MD,† Soo Hyun Song, PhD,‡ and Seung Wook Lee, MD, PhD†

*Department of Urology, Seoul National University Bundang Hospital, Seongnam, Korea; †Department of Urology, Hanyang University College of Medicine, Seoul, Korea; ‡Comparative Literature and Culture, Yonsei University, Underwood International College, Seoul, Korea

DOI: 10.1111/jsm.12784

ABSTRACT

Introduction. Quality assessment of randomized controlled trials (RCTs) is important to prevent the adoption of findings of low-quality trials into clinical practice.

Aim. The aim of this study was to analyze the quality of studies reporting RCTs in andrology journals (*The Journal of Sexual Medicine* [JSM], the *Asian Journal of Andrology* [AJA], the *Journal of Andrology* [JOA], the *International Journal of Andrology* [IJA]).

Methods. A quality assessment was conducted on all studies identified as RCTs published in andrology journals (*JSM*, *AJA*, *JOA*, *IJA*) until 2011. The review period was divided into three terms: early, mid, and late each journal.

Main Outcome Measures. The Jadad scale, van Tulder scale, and the Cochrane Collaboration Risk of Bias Tool (CCRBT) were employed. The RCTs were also categorized by country of origin, the inclusion of institutional review board (IRB) approval, funding, and blindness.

Results. There were1,954 original articles published in the *JSM*, 893 articles in the *AJA*, 2,527 articles in the *JOA*, and 2,086 articles in the *IJA* for the review period. There were 172 studies reporting on RCTs in the *JSM*, 33 RCTs in the *AJA*, 63 RCTs in the *JOA*, and 29 RCTs in the *IJA*. No significant increase in Jadad or van Tulder scale scores were found over time, nor were there any significant changes in the number of high-quality articles as assessed by CCRBT. However, significant differences in quality analysis were found according to blinding, funding, and IRB approval.

Conclusion. The number of original articles and RCTs in andrology increased over time. However, the ratio of RCTs to original articles as well as RCT quality was statistically insignificant. It would be required for the researchers to focus efforts in performing high-quality studies to ensure appropriate randomization, reviews by IRB, financial support, and inclusion of allocation concealment during study performance. **Jo JK, Chung JH, Kim KS, Song SH, and Lee SW. Reporting of randomized controlled trials in andrology journals: A quality assessment. J Sex Med 2015;12:350–357.**

Key Words. Risk; Quality; Randomized Controlled Trial

Introduction

Evidence-based medicine (EBM) is important to the use of current best evidence in making decisions for the diagnosis and treatment of disease. EBM is based on clinical practices based on the systemic review of studies on the causes, diagnosis, treatment, and prevention of a disease [1]. Randomized controlled trial (RCT) is the most reliable method to assess the effectiveness of a medical treatment as its study design has the least bias and the highest evidence level [2–4]. RCTs reduce the risk of bias during study design, advantage on the most valuable data among study methods, and then it can be the most reliable assessment of the effectiveness of medical treatments.

However, RCTs also have bias including the study design, the execution of the study, and the reporting of results, which can lead to an incorrect conclusion [5].

Objective assessment of articles on methodological quality can heighten the quality of medical care. Peer review within journals, which is an essential prevention method against error, can assess the validity of studies. And then it led to prevent incorrect information, which bias can be applied in clinic [6].

Complete assessment process can identify any unnecessary or erroneous data and it can eliminate incorrect clinical application due to incorrect information and save medical expenses [7].

A methodologic quality assessment of clinical trials can use individual scales. Scales on the quality assessment have the advantage of easy comparison between studies by quantitatively assessing the reporting quality.

The CONSORT statement is a generally accepted guideline for designing an RCT, and RCTs according to the CONSORT statement are considered the most appropriate for EBM. Further, the CONSORT statement has been applied in eminent journals [8–10]. However, the CONSORT statement is just a guideline for preparation of RCT, not a tool for quality assessment.

The Jadad scale, the van Tulder scale, and the Cochrane Collaboration Risk of Bias Tool (CCRBT) are representative tools for assessing the quality of an RCT. The Jadad scale has the advantage to assess quality of RCTs simply and easily and it contains three commonly used quality measures (randomization, double blinding, dropout), but it does not include an assessment for allocation concealment. Allocation concealment ensures the sequence in the study [11].

In addition, the van Tulder scale and the CCRBT are common tools that are used to assess RCT evidence and article validity. The van Tulder scale and the CCRBT do include an assessment item for allocation concealment.

Aims

The quality of RCTs published in andrology has not been analyzed to date, thus the aim of this study was to conduct a quality analysis of RCTs published in four representative andrology journals (*The Journal of Sexual Medicine* [JSM], *Asian Journal of Andrology* [AJA], *Journal of Andrology* [JOA], and *International Journal of Andrology* [IJA]) using representative tools for assessing the quality of RCT. Our aim was that this current study can help evaluate the factors involved in designing RCTs, and thereby help the quality of RCTs published in the field of andrology to continue to improve.

Methods

Study Cohort

A total of 1,954 original articles published in the *JSM* from 2004 to 2011 were searched using PubMed and Embase. A total of 893 articles published in *AJA* from 1999 (Vol. 1) to 2011 (Vol. 12), a total of 2,527 articles published in *JOA* from 1983 (Vol. 4) to 2011 (Vol.32), and a total of 2,086 articles published in *IJA* from 1980 (Vol. 3) to 2011 (Vol. 34) were also searched through a web search. When we have chosen representative journals of andrology, *JOA* and *IJA* just integrated, so we have analyzed four journals.

Selection of RCTs

Two independent reviewers determined all RCTs published in *JSM*, *AJA*, *JOA*, and *IJA*. Search terms such as "randomized," "randomization," and "randomly" were used. Two independent reviewers extracted the RCTs based on positive findings from the search terms. Only in any discrepancies between two reviewers a third reviewer judged equivocal points during scoring. Four journals have different years of publication, so we divided the periods into early, mid, and late terms each journal.

Assessment Tools

The quality assessment was conducted by the two reviewers using the Jadad scale, van Tulder scale, and CCRBT. If there was a difference in the assessment result, the result was adjusted again by the third reviewer.

Jadad Scale

The Jadad scale, which is also known as the Oxford quality scoring system, consists of five items used to assess RCT quality: two items related to randomization, two items related to blinding, and one item related to dropout. Considering randomization, one point is given if random assignment is cited in the RCT, one point is additionally given if an appropriate randomization method is used, and one point is subtracted for incorrect employment of randomization (range of available scores, 0–2). Considering blinding, one point is given if double

blinding is cited in the RCT, one point is additionally given if an appropriate double blinding method is used, and one point is subtracted for incorrect employment of double blinding (range of available scores, 0–2). One point is also given if dropout is cited in the RCT. Thus, RCT quality is assessed using a score out of five points. A score of two or less meant that the RCT was deemed to be of "low" quality, whereas a score greater than two meant that the RCT was deemed to be of "high" quality [3].

van Tulder Scale

The van Tulder scale has 11 items used to assess RCT quality: appropriateness of randomization, nondisclosure of treatment assignment, similarity of baseline characteristics, patient blinding, care provider blinding, investigator blinding, simultaneous arbitration, compliance, dropout/dropout rate, the time interval at which results were assessed, and random assignment. Each item is answered "yes," "no," or "don't know." A score greater than four points meant that the RCT was deemed to be of "high" quality [12].

CCRBT

The CCRBT has six items used to assess RCT quality: sequence determination, nondisclosure of assignment, blinding, incomplete end points, reporting selective results, and other potential bias threatening feasibility. Each item is assessed with "yes," "no," or "unclear," referring to the risk of bias associated with the item. Overall, RCT quality is classified as low, moderate, or high risk of bias, depending on how the items were answered. A "yes" answer for the first three items and a "no" answer for the last three items mean that the RCT is deemed to have a "low" risk of bias. If two or less items are answered as "no" or "unclear," then the RCT is deemed to have a "moderate" risk of bias. If three or more items are answered as "no" or "unclear," then the RCT is deemed to have a "high" risk of bias [13].

Analysis of Other Characteristics of RCTs

RCTs were also categorized by country of origin, topic, institutional review board (IRB) approval, blinding, and funding. Differences in quality for each of these parameters were calculated.

Statistical Analysis

The Kruskal–Wallis test and one-way analysis of variance were used for comparison of quality scores for the Jadad and Van Tulder scale. The chi-square test was used for the number of RCT distribution by country origin and topics. Logistic regression analyses were performed to analyze the relation of each variable with quality of article. SPSS v.19.0 (IBM Corp., Armonk, NY, USA) was used for all statistical analyses. A P value less than 0.05 was considered statistically significant.

Main Outcome Measures

Quantitative Variation of RCTs over Time

A total of 1,954 original articles published in the *JSM*, from 2004 to 2011, were searched, of which 172 were deemed to be RCTs. There were 39 (14.89%), 70 (7.77%), and 63 (7.96%) RCTs published in the early, mid, and late terms, respectively ($P = 0.001$) A total of 893, 2,527, and 2,086 original articles published in *AJA*, *JOA*, and *IJA* were also searched, respectively (Table 1).

Qualitative Variation of RCTs over Time

Jadad Quality Assessment Scale

The mean Jadad scale scores were 2.18 ± 0.82, 2.40 ± 1.08, and 2.52 ± 1.24 for the early, mid, and late terms in *JSM*, respectively ($P = 0.008$). The number of high quality articles was 13 (33.33%), 28 (40%), and 28 (44.44%) for the early, mid, and late terms in *JSM*, respectively ($P = 0.538$). In addition, the mean Jadad scale scores were 1.38 ± 1.19, 2.09 ± 1.22, and 2.21 ± 1.37 in *AJA*; 2.00, 2.56 ± 0.73, and 2.10 ± 0.91 in *JOA*; and 2.20 ± 1.14, 1.67 ± 0.52, and 1.77 ± 0.83 in *IJA* for the early, mid, and late terms, respectively ($P = 0.679$, $P = 0.073$, $P = 0.220$). The number of high-quality articles was 1 (12.5%), 3 (27.27%), and 5 (35.71%) in *AJA*; 0, 6 (66.67%), and 10 (32.26%) in *JOA*; and 3 (30%), 0, and 3 (23.08%) in *IJA* for the early, mid, and late terms, respectively ($P = 0.501$, $P = 0.066$, $P = 0.343$) (Table 2).

van Tulder Assessment Scale

The mean van Tulder scale scores were 5.28 ± 1.56, 5.47 ± 1.96, and 5.56 ± 1.74 in *JSM* for the early, mid, and late terms, respectively ($P = 0.877$). The number of high-quality articles was 26 (66.67%), 49 (70%), and 45 (71.43%) in *JSM* for the early, mid, and late terms, respectively ($P = 0.877$). In addition, the mean van Tulder scale scores were 2.63 ± 1.41, 4.36 ± 1.80, and 4.79 ± 2.19 in *AJA*; 4.33 ± 1.16, 4.78 ± 0.97, and 4.48 ± 1.61 in *JOA*; and 3.50 ± 2.22, 3 ± 1.41,

Table 1 Characteristics of RCTs according to publication year

Journals	Years	Original articles	RCT (%)	IRB (%)	Blinding (%)	Concealment of allocation (%)	Funding (%)
The Journal of Sexual Medicine	2004–2006	262	39 (14.89)	29 (74.36)	27 (69.23)	1 (2.56)	18 (46.15)
	2007–2009	901	70 (7.77)	59 (84.29)	52 (74.29)	5 (7.14)	50 (71.43)
	2010–2011	791	63 (7.96)	47 (74.60)	42 (66.67)	4 (6.35)	36 (57.14)
	P value		0.001	0.309	0.621	0.603	0.028
	Total	1,954	172 (8.80)	135 (78.49)	121 (70.35)	10	104 (60.47)
Asian Journal of Andrology	1999–2003	212	8 (3.77)	2 (25)	2 (25)	0	0
	2004–2007	276	11 (3.99)	6 (54.55)	5 (45.45)	0	4 (36.36)
	2008–2011	405	14 (3.46)	12 (85.71)	7 (50)	0	6 (42.86)
	P value		0.935	0.017	0.505		0.095
	Total	893	33 (3.70)	20 (60.61)	14 (42.42)	0	10 (30.30)
Journal of Andrology	1983–1992	589	3 (0.51)	2 (66.67)	3 (100)	0	1 (33.33)
	1993–2002	971	9 (0.93)	7 (77.78)	7 (77.78)	0	5 (55.56)
	2003–2011	967	31 (3.21)	25 (80.65)	13 (41.94)	1 (3.23)	22 (70.97)
	P value		<0.001	0.846	0.041	0.82	0.339
	Total	2,527	43 (1.70)	34 (79.07)	23 (53.49)	1 (2.33)	28 (65.12)
International Journal of Andrology	1980–1990	676	10 (1.48)	1 (10)	3 (30)	2 (20)	2 (20)
	1991–2001	640	6 (0.94)	4 (66.67)	2 (33.33)	1 (16.67)	1 (16.67)
	2002–2011	770	13 (1.69)	10 (76.92)	4 (30.77)	0	4 (30.77)
	P value		0.474	0.004	0.99	0.251	0.745
	Total	2,086	29 (1.39)	15 (51.72)	9 (31.03)	3 (10.34)	7 (24.14)

Chi-square test
IRB = institutional review board; RCT = randomized controlled trial

and 4 ± 1.47 in *IJA* for the early, mid, and late terms, respectively ($P = 0.074$, $P = 0347$, $P = 0.281$). The number of high-quality articles was 1 (12.5%), 5 (45.45%), and 7 (50%) in *AJA*; 2 (66.67), 6 (66.67%), and 13 (41.94%) in *JOA*; and 2 (20%), 1 (16.67%), and 6 (46.15%) in *IJA* for the early, mid, and late terms, respectively ($P = 0.197$, $P = 0.347$, $P = 0.281$) (Table 2).

Table 2 Quality assessment of RCTs according to publication year

		Jadad scale		van Tulder scale		Cochrane's assessment of risk bias		
Journals	Years	Score	High quality (%)	Score	High quality (%)	High risk (%)	Moderate risk (%)	Low risk (%)
The Journal of Sexual Medicine	2004–2006	2.18 ± 0.82	13 (33.33)	5.28 ± 1.56	26 (66.67)	11 (28.20)	28 (71.79)	0
	2007–2009	2.40 ± 1.08	28 (40)	5.47 ± 1.96	49 (70)	21 (30)	45 (64.29)	4 (5.71)
	2010–2011	2.52 ± 1.24	28 (44.44)	5.56 ± 1.74	45 (71.43)	21 (33.33)	40 (63.49)	2 (3.17)
	P value	0.008	0.538*	0.327	0.877*	0.576*		
Asian Journal of Andrology	1999–2003	1.38 ± 1.19	1 (12.5)	2.63 ± 1.41	1 (12.5)	7 (87.5)	1 (12.5)	0
	2004–2007	2.09 ± 1.22	3 (27.27)	4.36 ± 1.80	5 (45.45)	7 (63.64)	4 (36.36)	0
	2008–2011	2.21 ± 1.37	5 (35.71)	4.79 ± 2.19	7 (50)	7 (50)	7 (50)	0
	P value	0.679†	0.501*	0.074†	0.197*	0.213*		
Journal of Andrology	1983–1992	2	0	4.33 ± 1.16	2 (66.67)	3 (100)	0	0
	1993–2002	2.56 ± 0.73	6 (66.67)	4.78 ± 0.97	6 (66.67)	9 (100)	0	0
	2003–2011	2.10 ± 0.91	10 (32.26)	4.48 ± 1.61	13 (41.94)	30 (96.77)	1 (3.23)	0
	P value	0.073	0.066*	0.127	0.347*	0.820*		
International Journal of Andrology	1980–1990	2.20 ± 1.14	3 (30)	3.50 ± 2.22	2 (20)	8 (80)	2 (20)	0
	1991–2001	1.67 ± 0.52	0	3.00 ± 1.41	1 (16.67)	6 (100)	0	0
	2002–2011	1.77 ± 0.83	3 (23.08)	4.00 ± 1.47	6 (46.15)	13 (100)	0	0
	P value	0.22	0.343*	0.356	0.281*	0.130*		
The Journal of Sexual Medicine	Total	2.40 ± 1.10	69 (40.12)	5.46 ± 1.79	120 (69.77)	53 (30.81)	113 (65.69)	6 (3.49)
Asian Journal of Andrology	Total	1.97 ± 1.29	9 (27.27)	4.12 ± 2.04	13 (39.39)	21 (63.64)	12 (36.36)	0
Journal of Andrology	Total	2.19 ± 0.85	16 (37.21)	4.53 ± 1.45	21 (48.84)	42 (92.67)	1 (2.33)	0
International Journal of Andrology	Total	1.90 ± 0.90	6 (20.69)	3.62 ± 1.74	9 (31.03)	27 (93.10)	2 (6.90)	0
	P value	0.057	0.149*	0.095	<0.001*	<0.001*		

One-way ANOVA
*Chi-square test
†Kruskal–Wallis test
ANOVA = analysis of variance; RCT = randomized controlled trial

Table 3 Distribution of country, (%)

Journal name Country of origin	*The Journal of Sexual Medicine*	*Asian Journal of Andrology*	*Journal of Andrology*	*International Journal of Andrology*	Total	*P* value
United States	51 (65.38)	0	25 (32.05)	2 (2.56)	78 (28.16)	0.195
Italy	14 (50)	3 (10.71)	5 (17.86)	6 (21.43)	28 (10.11)	
United Kingdom	9 (60)	0	4 (26.67)	2 (13.33)	15 (5.42)	
China	2 (13.33)	12 (80)	0	1 (6.67)	15 (5.42)	
Germany	8 (57.14)	2 (14.29)	2 (14.29)	2 (14.29)	14 (5.05)	
The Netherlands	11 (78.57)	0	2 (14.29)	1 (7.14)	14 (5.05)	
Korea	9 (75)	3 (25)	0	0	12 (4.33)	
Canada	10 (100)	0	0	0	10 (3.61)	
Other	58 (63.74)	13 (14.29)	5 (5.49)	15 (16.48)	91 (32.85)	

Chi-square test

CCRBT

No RCTs with low risk of bias were found in *AJA*, *JOA*, and *IJA* using the CCRBT. The number of articles with low risk of bias was 0, 4 (5.71%), and 2 (3.17%) in *JSM* for the early, mid, and late terms, respectively. The number of articles with moderate risk of bias was 28 (71.79%), 45 (64.29%), and 40 (63.49%) in *JSM*; 1 (12.5%), 4 (36.36%), and 7 (50%) in *AJA*; 0, 0, and 1 (3.23%) in *JOA*; and 2 (20%), 0, and 0 in *IJA* for the early, mid, and late terms, respectively. The number of articles with high risk of bias was 11 (28.20%), 21 (30%), and 21 (33.33%) in *JSM*; 7 (87.5%), 7 (63.64%), and 7 (50%) in *AJA*; 3 (100%), 9 (100%), and 30 (96.77%) in *JOA*; and 8 (80%), 6 (100%), and 13 (100%) in *IJA* for the early, mid, and late terms, respectively ($P = 0.576$, $P = 0.213$, $P = 0.820$, $P = 0.130$; Table 2).

Results

Analysis of Factors Related to the Quality of the Articles

No significant difference in the number of RCT distribution by country of origin was found ($P = 0.212$; Table 3). According to the number of RCT distribution by topic, there were shown significant difference ($P < 0.001$; Table 4). A significant difference in the number of RCTs approved by an IRB was found over time in *AJA* and *IJA* ($P = 0.017$, $P = 0.004$). In *JSM* and *JOA*, the number of RCTs approved by an IRB was not shown as statistically significant ($P = 0.309$, $P = 0.846$). In addition, there was a significant correlation between RCT quality and IRB approval by both the Jadad scale and the van Tulder scale in RCTs in the four journals ($P = 0.001$, $P < 0.001$). No quantitative difference in funding was found over time except in *JSM* ($P = 0.095$, $P = 0.339$, $P = 0.745$). The van Tulder scale showed a significant correlation between funding and high RCT quality in *JSM*, *AJA*, and in the four journals combined ($P = 0.003$). No quantitative change of RCT was found for blinding over time except in *JOA* ($P = 0.505$, $P = 0.621$, $P = 0.99$), but there were a significantly greater number of high-quality articles for RCTs with blinding by all three scales in the four journals combined ($P < 0.001$). In multivariable logistic regression, blinding is the most

Table 4 Distribution of topic, (%)

		Journal name					
		The Journal of Sexual Medicine	*Asian Journal of Andrology*	*Journal of Andrology*	*International Journal of Andrology*	Total	*P* value
Topic	Couples sex	6 (3.5)	0	0	0	6 (2.2)	<0.001
	ED	95 (55.2)	15 (45.5)	7 (16.3)	4 (13.8)	121 (43.7)	
	Ejaculatory disorder	12 (7.0)	0	0	0	12 (4.3)	
	FSD	33 (19.2)	0	0	0	33 (11.9)	
	Hormone	1 (0.6)	1 (3.0)	15 (34.9)	9 (31)	26 (9.4)	
	Hypogonadism	5 (2.9)	0	1 (2.3)	1 (3.4)	7 (2.5)	
	Infertility	0	4 (12.1)	2 (4.7)	6 (20.7)	12 (4.3)	
	Peyronie's disease	4 (2.3)	0	1 (2.3)	0	5 (1.8)	
	Prostate	3 (1.7)	8 (24.2)	4 (9.3)	0	15 (5.4)	
	Semen	0	0	6 (14.0)	2 (6.9)	8 (2.9)	
	Sexual function	8 (4.7)	0	0	0	8 (2.9)	
	Other	5 (2.9)	5 (15.2)	7 (16.3)	7 (24.1)	24 (8.7)	

ED = erectile dysfunction; FSD = female sexual disorder

Table 5 Logistic regression analysis of related factor with quality of article

	Jadad risk (number of high quality)		van Tulder risk (number of high quality)	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Period				
Early	Ref.		Ref.	
Mid	1.398 (0.665–2.939)	0.376	0.964 (0.448–2.076)	0.926
Late	1.564 (0.754–3.244)	0.23	0.736 (0.345–1.569)	0.736
IRB				
None	Ref.		Ref.	
With IRB	1.941 (1.014–3.714)	0.045	2.767 (1.476–5.188)	0.002
Funding				
None	Ref.		Ref.	
With funding	0.89 (0.518–1.527)	0.672	2.671 (1.526–4.675)	0.001
Blinding				
None	Ref.		Ref.	
With blinding	3.005 (1.665–5.424)	<0.001	4.351 (2.481–7.629)	<0.001

CI = confidential interval; IRB = institutional review board; OR = odds ratio

powerful variable in Jadad risk and van Tulder risk ($P < 0.001$; Table 5).

Discussion

No significant differences in the quality of RCTs published in the four journals were detected over time. The number of high-quality articles increased when IRB approval, blinding, and funding had been applied.

Since the introduction of EBM, preexisting knowledge of diseases and treatment options has been refined through the introduction of stronger medical evidence. The majority of EBM studies have been conducted in Canada and England. In particular, medical RCTs have been collated and assembled in a database by the Cochrane Collaboration with the help of clinical epidemiologists worldwide. These data have particularly helped to establish EBM, guidelines from which are now being widely used by clinicians [14].

Moher et al. reported comparative before-and-after evaluation in which reports of RCTs published in 1994 were compared with RCT reports from the same journals published in 1998. They included 211 reports from *British Medical Journal* (BMJ), *The Journal of the American Medical Association* (JAMA), and *The Lancet* (journals that adopted CONSORT) as well as *The New England Journal of Medicine*. They showed the number of RCTs increased in all four journals in 1998 compared with 1994. But the quantitative status of the RCTs analyzed here did not significantly change over time [15].

Quality assessment of RCTs enables bias to be detected, which may occur during the study design, execution, and analysis phases [16,17]. Quality assessment is important as a means of determining whether the results of a study are acceptable and whether additional studies may be required for result validation.

Zhang et al. conducted a quality assessment of RCTs published in China. They used CONSORT statement as a tool for quality assessment of RCTs [18]. Xu et al. also used the CONSORT statement to conduct a quality assessment of RCTs published in major Chinese journals [19]. However, as the CONSORT statement is a set of guidelines rather than a tool for quality assessment, it cannot compare and analyze the quality of studies quantitatively. Validated quality assessment tools for RCTs include the Jadad scale, the van Tulder scale, the CCRBT, Newell's scale, the Scottish Intercollegiate Guidelines Network, and the National Institute for Health and Clinical Excellence guidelines. We chose three tools that enabled comprehensive analysis of the various items contained within the CONSORT statement [20]. Compared with the Jadad scale, the van Tulder scale is more effective in RCT quality assessment, as it includes the majority of assessment items contained within the CONSORT statement. However, no standardized tool for RCT quality assessment is currently available. The use of three assessment tools in the present study enabled broad consideration of the CONSORT statement guidelines and also provided a higher level of objectivity in the assessment of RCT quality [21,22].

Hewitt et al. conducted an assessment of RCTs published in four high-impact journals (*BMJ*, *JAMA*, *The Lancet*, and *The New England Journal of Medicine*) and reported that the frequency of conducting concealment of allocation inappropriately or uncertainly was 46% [23]. If concealment of allocation is conducted inappropriately, it could

negatively affect randomization during clinical studies, which is likely to affect the quality of reporting. We found that among the RCTs published in andrology, no articles correctly described concealment of allocation in *AJA*, but other journals described concealment of allocation. Thus, correct implementation and description of concealment of allocation may lead to an overall improvement in RCT quality.

Articles approved by an IRB had higher quality scores, and there was a greater number of high-quality articles. IRB approval is given after a positive assessment of the design and feasibility of a clinical study, and it is considered an international standard; IRB approval is an important step in ensuring the quality of an RCT.

Clifford et al. analyzed 100 RCTs published in five peer-reviewed, with high impact factor, general medical journals and reported that there was no correlation between the study quality and the funding source [24]. In contrast, we found, using the van Tulder scale and CCRBT, that the number of high-quality articles was significantly higher for articles that had received funding. We also found a significant difference in RCT quality for blinding, which was noted using all three tools. This highlights the importance of using blinding to enhance the quality of an RCT.

Moreover, there are some tools that can assess to quality of non-RCT [25–28]. Wang et al. assessed the quality of RCT and non-RCT [29]. In the present study, we focused on quality of RCT.

The results presented here should be interpreted within the study limitations. First, no one representative assessment tool is available for the qualitative analysis of an RCT. Also, there is not one tool that can assess all of the items listed in the CONSORT statement. However, three representatively used tools for quality assessment of RCTs that covered the majority of items within the CONSORT statement were used to supplement this limitation. Second, because of the nature of the manual searching and evaluation used in this study, assessor bias may have influenced the selection and/or assessment process. This limitation was minimized through the use of two reviewers who independently extracted and assessed the RCTs, as well as the use of a third reviewer who moderated any discrepancies.

Conclusion

The numbers of original articles and RCTs published in andrology has increased over time. However, no significant increase in the ratio of RCTs to original articles and in RCT quality was found. The number of high-quality articles increased when IRB approval had been granted, there was blinding, or there was funding provided. Researchers should focus efforts in performing high-quality studies.

Corresponding Author: Seung Wook Lee, MD, Department of Urology, Hanyang University Guri Hospital, 249-1, Gyomun-dong, Guri 471-701, Korea. Tel: +82-31-560-2374; Fax: +82-31-560-2372; E-mail: swleepark@hanyang.ac.kr

Conflict of Interest: The authors report no conflicts of interest.

References

1 Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: What it is and what it isn't. 1996. Clin Orthop Relat Res 2007;455:3–5.
2 Uetani K, Nakayama T, Ikai H, Yonemoto N, Moher D. Quality of reports on randomized controlled trials conducted in Japan: Evaluation of adherence to the CONSORT statement. Intern Med 2009;48:307–13.
3 Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Control Clin Trials 1996;17:1–12.
4 Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 2001;323:42–6.
5 Featherstone K, Donovan JL. Random allocation or allocation at random? Patients' perspectives of participation in a randomised controlled trial. BMJ 1998;317:1177–80.
6 Jackson JL, Srinivasan M, Rea J, Fletcher KE, Kravitz RL. The validity of peer review in a general medicine journal. PLoS ONE 2011;6:e22475.
7 Lim SM, Shin ES, Lee SH, Seo KH, Jung YM, Jang JE. Tools for assessing quality and risk of bias by levels of evidence. J Korean Med Assoc AID 2011;54:419–29.
8 Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637–9.
9 Altman DG. Endorsement of the CONSORT statement by high impact medical journals: Survey of instructions for authors. BMJ 2005;330:1056–7.
10 Gagnier J, Boon H, Rochon P, Barnes J, Moher D, Bombardier C. Improving the quality of reporting of randomized controlled trials evaluating herbal interventions: Implementing the CONSORT statement [corrected]. Explore (NY) 2006;2:143–9.
11 Clark HD, Wells GA, Huet C, McAlister FA, Salmi LR, Fergusson D, Laupacis A. Assessing the quality of randomized trials: Reliability of the Jadad scale. Control Clin Trials 1999;20:448–52.
12 van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the Cochrane collaboration back review group. Spine 2003;28:1290–9.
13 Higgins JPT, ed. Cochrane handbook for systematic reviews of interventions. Vol. 5. Chichester: Wiley-Blackwell; 2008.
14 White PJ. Evidence-based medicine for consumers: A role for the Cochrane Collaboration. JMLA 2002;90:218–22.

15 Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: A comparative before-and-after evaluation. JAMA 2001;285: 1992–5.
16 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408–12.
17 Chalmers TC, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. N Engl J Med 1983;309:1358–61.
18 Zhang D, Yin P, Freemantle N, Jordan R, Zhong N, Cheng KK. An assessment of the quality of randomised controlled trials conducted in China. Trials 2008;9:22.
19 Xu L, Li J, Zhang M, Ai C, Wang L. Chinese authors do need CONSORT: Reporting quality assessment for five leading Chinese medical journals. Contemp Clin Trials 2008;29:727–31.
20 Kim SW, Choi YS, Ahn HS, Lee HY, Ahn DS, Lee YM. Quantity and quality assessment of randomized controlled trials published in five Korean medical journals, from 1980 to 2000. J Korean Acad Fam Med 2004;25:118–25.
21 Moher D, Schulz KF, Altman DG. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. Clin Oral Investig 2003;7:2–7.
22 Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. Control Clin Trials 1995;16:62–73.
23 Hewitt C, Hahn S, Torgerson DJ, Watson J, Bland JM. Adequacy and reporting of allocation concealment: Review of recent trials published in four general medical journals. BMJ 2005;330:1057–8.
24 Clifford TJ, Barrowman NJ, Moher D. Funding source, trial outcome and reporting quality: Are they related? Results of a pilot study. BMC Health Serv Res 2002;2:18.
25 Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998;52:377–84.
26 Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa: Ottawa Hospital Research Institute; 2000. Available at: http://www.ohri.ca/prog rams/clinical_epidemiology/oxford .asp (accessed March 28, 2011). [Internet].
27 Reisch JS, Tyson JE, Mize SG. Aid to the evaluation of therapeutic studies. Pediatrics 1989;84:815–27.
28 Thomas H. Quality assessment tool for quantitative studies: Effective public health practice project. Hamilton: McMaster University; 2003.
29 Wang YL, Chen Y, Wang Y, Wang SM. Assessment of the quality of clinical therapeutic research published in Chinese Journal of Pediatrics. Zhonghua Er Ke Za Zhi 2013;51:29–33, (in Chinese).