

## Original Article

# Assessments of the quality of randomized controlled trials published in *International Journal of Urology* from 1994 to 2011

Hee Ju Cho,<sup>1</sup> Jae Hoon Chung,<sup>2</sup> Jung Ki Jo,<sup>2</sup> Dong Hyuk Kang,<sup>2</sup> Jeong Man Cho,<sup>1</sup> Tag Keun Yoo<sup>1</sup> and Seung Wook Lee<sup>2</sup>

<sup>1</sup>Department of Urology, Eulji University School of Medicine, and <sup>2</sup>Department of Urology, Hanyang University College of Medicine, Seoul, Korea

**Abbreviations & Acronyms**

CCRB = Cochrane Collaboration Risk of Bias Tool  
CONSORT = Consolidated Standards of Reporting Trials  
EBM = evidence-based medicine  
IF = impact factor  
IJU = *International Journal of Urology*  
IRB = institutional review board  
KJFM = *Korean Journal of Family Medicine*  
KJU = *Korean Journal of Urology*  
RCT = randomized controlled trials

**Correspondence:** Seung Wook Lee M.D., Department of Urology, Hanyang University Guri Hospital, 249-1 Gyomun-dong, Guri 471-701, Korea. Email: swleepark@hanyang.ac.kr

Received 11 July 2012; accepted 5 March 2013.

**Objectives:** Randomized controlled trials are one of the most reliable resources for assessing the effectiveness and safety of medical treatments. Low quality randomized controlled trials carry a large bias that can ultimately impair the reliability of their conclusions. The present study aimed to evaluate the quality of randomized controlled trials published in *International Journal of Urology* by using multiple quality assessment tools.

**Methods:** Randomized controlled trials articles published in *International Journal of Urology* were found using the PubMed MEDLINE database, and qualitative analysis was carried out with three distinct assessment tools: the Jadad scale, the van Tulder scale and the Cochrane Collaboration Risk of Bias Tool. The quality of randomized controlled trials was analyzed by publication year, type of subjects, intervention, presence of funding and whether an institutional review board reviewed the study.

**Results:** A total of 68 randomized controlled trial articles were published among a total of 1399 original articles in *International Journal of Urology*. Among these randomized controlled trials, 10 (2.70%) were from 1994 to 1999, 23 (4.10%) were from 2000 to 2005 and 35 (4.00%) were from 2006 to 2011 ( $P = 0.494$ ). On the assessment with the Jadad and van Tulder scale, the numbers and percentage of high quality randomized controlled trials increased over time. The studies that had institutional review board reviews, funding resources or that were carried out in multiple institutions had an increased percentage of high quality articles.

**Conclusions:** The numbers and percentage of high-quality randomized controlled trials published in *International Journal of Urology* have increased over time. Furthermore, randomized controlled trials with funding resources, institutional review board reviews or carried out in multiple institutions have been found to be of higher quality compared with others not presenting these features.

**Key words:** evidence-based medicine, Japan, quality assessment, randomized controlled trial, urology.

## Introduction

Since the study of EBM was initially introduced by the EBM Working Group in 1992, EBM has become the fundamental method for drawing conclusions in modern medicine. Usage of RCT minimizes the biases that can occur during research. For this reason, RCT are thought to provide the most dependable evidence in EBM research, which itself is one of the most reliable research methods for assessing the effectiveness and safety of medical treatments.<sup>1</sup> However, with low quality RCT resulting from improper basic research instruments, such as randomization, blinding and allocation concealment, the reliability of conclusions that are drawn can be impaired.<sup>2</sup> Therefore, the assessments of the quality of RCT are critical for raising the reliability of EBM. The CONSORT statement was announced in 1996 with an aim to decrease the number of poorly conducted RCT by offering appropriate standards. Since the CONSORT statement, various journals and organizations have played a guiding role in enhancing the quality of RCT.<sup>3</sup> The CONSORT statement is a guideline for writing

RCT, and several studies have used this to evaluate the quality of RCT. These studies used the conformity to allocate arbitrary scores for comparison; however, as this is not a scoring tool, we used the CCRBT in the present study. By contrast, the Jadad scale offers a scoring system to independently assess RCT. The Jadad scale assesses scale-based research through three simple and easy items: randomization, double blinding and dropouts. Although the ease and simplicity of the Jadad scale has allowed it to become widely used, it does not include an assessment item for allocation concealment.<sup>4</sup> Unlike the Jadad scale, both the van Tulder scale and CCRBT assess allocation concealment. Allocation concealment is one method of carrying out a randomized allocation sequence, and is essential to avoid selection bias in patient allocations before treatment. The van Tulder scale is one of the tools recommended by the Cochrane group and is composed of 11 items. When an article corresponds with the conditions of five of these 11 items, then it is rated as high quality.<sup>5</sup> CCRBT assesses the quality of RCT using six classifications, and RCT are ultimately categorized as high, moderate or low risk of bias.<sup>6</sup> IJU is the official journal of the Japanese Urological Association and has published evidence-based and scientifically written articles since 1994. However, no study has yet assessed the quality of RCT published in IJU. The present study aims to evaluate the quality of RCT published in IJU by using quality assessment tools for RCT including the Jadad scale, the van Tulder scale and CCRBT.

## Methods

### Subjects

A total of 1399 original articles published in IJU over the past 18 years from 1994 (volume 1) to 2011 (volume 18) were manually searched.

### Determination of RCT

Two reviewers found RCT reports published in IJU independently by using the PubMed MEDLINE database. They found and collected RCT reports by searching for keywords, such as “random”, “randomized” and “randomly”, in the Methods section of the articles. A third reviewer then made a final determination by adjusting the collected data.

### Assessment of the quality of RCT

Quality assessment was carried out using the Jadad scale, the van Tulder scale and CCRBT. Quality was assessed as an individual index. Two reviewers carried out the assessments, and if there were discrepancies in the results, then these were resolved through discussion. We blinded the authors to ensure that any subjective opinions of the evaluator based on their affiliations or connections with the authors did not affect their assessment of the quality of the RCT. The quality assessments were carried out by determining the publication

year, type of subjects, interventions, presence of funding, whether an IRB reviewed the trial and whether the trial was carried out as a multicenter study.

### Jadad scale

The Jadad scale, also known as the Oxford quality scoring system, is a scale-scoring tool that was developed in 1996 through a standardized item reduction process of 49 items. It shows a high level of interrater agreement, with an interclass correlation coefficient of 0.66 (95% CI 0.53–0.79) and high validity.<sup>7</sup> It is composed of five total points: two points that are related to randomization, two points that are related to blinding and one point that is related to dropouts.<sup>7</sup> In the analysis carried out in the present study, when the trial only provided comments without any descriptions of randomization and blinding, one point was granted for each item. One point was then added if there was any description of the proper methods. In contrast, when the description method was inappropriate, one point was deducted from each item. As for dropouts, when the number of dropouts in each subject group and the reasons for dropouts was specified, one point was given. If there were no dropouts, then it needed to be explicitly stated. When the total was  $\geq 3$  points, then the trial was assessed to be high quality. However, when the total was  $\leq 2$  points, then the trial was assessed to be low quality. For studies where double blinding was impossible, if the total score was  $\geq 2$  points, then the trial was assessed to be high quality.

### van Tulder scale

The van Tulder scale is a scale tool that was developed from the “Maastricht-Amsterdam LIST” (19 items), which itself was developed by the addition of a few items to the Delphi List that was used for the analysis of bias. The scale was revised to contain 11 components in 2003 and has been recommended by the Cochrane Collaboration Back Review Group for the methodological assessment of RCT.<sup>5</sup> Several studies have proven the mutual agreement between the van Tulder scale and other analysis tools, and the van Tulder scale has been used as a quality scoring method.<sup>8–11</sup> The van Tulder scale is designed to assess 11 components, including randomization, allocation concealment, baseline characteristics, patient blinding, caregiver blinding, observer blinding, co-intervention, compliance, dropout rate, end-point and intention to treat analysis.<sup>5</sup> In the analysis carried out in the present study, the scoring was performed by assigning “yes”, “no” or “don’t know” for each item, and if a minimum of five of the criteria were satisfied ( $\geq 5$  points), then the trial was assessed to be high quality.

### CCRBT

The CCRBT is an evaluation tool recommended by the Cochrane collaboration and carried out systemic analyses including meta-analyses. Unlike the Jadad and van Tulder

**Table 1** Characteristics of RCTs according to publication year

Years	Original articles	RCT (%)	Intervention: drug	Double blinding (%)	Concealment of allocation (%)
1994–1999	370	10 (2.70)	7 (70)	0	0
2000–2005	566	23 (4.1)	14 (60.9)	3 (13.0)	1 (4.3)
2006–2011	879	35 (4.00)	27 (77.1)	8 (22.9)	5 (14.3)
<i>P</i> -value		0.494	0.412	0.197	0.242
Total	1399	68 (3.7)	48 (70.6)	11 (16.2)	6 (8.8)

 $\chi^2$ -test.

scales, it cannot provide evaluation scoring. Instead, it uses six “sources of risk of bias”, including selection bias, performance bias, detection bias, attrition bias, reporting bias and other bias, for qualitative assessment based on the criteria for judgment. This allows RCT to be classified as “high risk”, “moderate risk” or “low risk” of bias based on their qualities.<sup>12</sup> In the analysis carried out in the present study, the scoring was performed by assigning “yes”, “no” or “unclear” for each item, where “yes” meant a low risk of bias, “no” meant a high risk of bias and “unclear” meant an uncertain risk of bias. If the first three questions were answered with “yes”, and no important concerns related to the last three domains were identified, then the trial was classified to be at low risk of bias. If the case was assessed in  $\leq 2$  domains with “unclear” or “no”, then the trial was classified to be at moderate risk of bias. If the case was assessed in  $\geq 3$  domains with “unclear” or “no”, then the trial was classified to be at high risk of bias.

## Statistical analysis

Starting with 1994, the qualities of the RCT were analyzed in 6-year units. Kruskal–Wallis analysis of variance was used to compare and analyze the respective scores obtained by each of the assessment tools. The  $\chi^2$ -test was used to compare and analyze the ratio of the high quality dissertations, and the quality assessment outcomes when CCRBT was used. SPSS version 18.0 (SPSS, Chicago, IL, USA) was used for statistical analysis, and a *P*-value  $< 0.05$  was considered to be statistically significant.

## Results

### Quantitative variation in the number of RCT published in IJU over time

From 1994 to 2011, 68 of the 1399 articles published in IJU were RCT. Among these 68 articles, 10 (2.70%) were published from 1994 to 1999, 23 (4.10%) were published from 2000 to 2005 and 35 (4.00%) were published from 2006 to 2011. Although the numbers of both published original articles and RCT have increased over time, the percentage of RCT in IJU has not significantly changed (*P* = 0.494; Table 1).

### Qualitative variation in the number of RCT published in IJU over time

- 1 Jadad quality assessment scale: When the RCT were assessed starting in 1994 using 6-year units, the mean Jadad scale score of RCT from 1994 to 1999 was  $1.20 \pm 0.42$ . From 2006 to 2011, the mean Jadad scale score significantly increased to  $2.48 \pm 1.38$  (*P* = 0.004). Although there were no high-quality articles from 1994 to 1999, there were 19 out of 35 (54.3%) high quality RCT from 2006 to 2011 (*P* = 0.005).
- 2 van Tulder assessment scale: The mean van Tulder scale score of RCT published in IJU from 1994 to 1999 was  $2.40 \pm 0.70$ , whereas the score of RCT published from 2006 to 2011 was  $5.54 \pm 2.57$  (*P* < 0.001). Although only one out of 10 (10%) high-quality RCT was published from 1994 to 1999, 19 out of 35 (54.3%) RCT published from 2006 to 2011 were of high quality (*P* = 0.022).
- 3 CCRBT: Among all RCT published in IJU from 1994 to 2005, no articles were assessed as at low risk of bias using CCRBT. However, from 2006 to 2011 there were five (14.3%) low risk of bias articles (*P* = 0.033) (Table 2).
- 4 Before and after 2001, which is when IF scores were introduced: The Jadad and van Tulder scores of 13 papers published before 2001 were  $1.31 \pm 0.48$  and  $2.27 \pm 1.01$ , respectively. The proportion of high-quality articles published before 2001 according to the Jadad and van Tulder scores were 1/17 (7.7%) and 2/13 (15.4%), respectively. For the 55 RCTs published after 2001, the Jadad and van Tulder scores increased to  $2.20 \pm 1.28$  and  $5.02 \pm 2.47$ , respectively. The proportion of high-quality articles published after 2001 according to the Jadad and van Tulder scores increased to 26/55 (47.3%) and 25/55 (45.5%), respectively (*P* < 0.005). For analyses using CCRBT, there were no publications with low risk of bias before 2001, whereas five publications published after 2001 were in this category (Table 3).

### Analysis of RCT by subject

Among all of the RCT published in IJU over the past 18 years, 22 were about voiding dysfunction, 18 were about oncology, 10 were about stone disease/endourology/laparoscopic surgery, and seven were about infertility/andrology. When the Jadad and van Tulder scales were used to assess the quality of each subject, the mean scores of RCT about voiding dysfunction and infertility/andrology were significantly higher than the other subjects (*P* = 0.004). When CCRBT was used to assess quality, two articles about voiding dysfunction, and one article about each of stone disease, infection and andrology were evaluated to be at low risk of bias. The differences in qualities of RCT according to each subject were statistically significant (Table 4).

**Table 2** Quality assessment of RCT according to publication year

Years	Jadad scale		van Tulder scale		Cochrane's assessment of risk bias		
	Score	High quality (%)	Score	High quality (%)	High risk (%)	Moderate risk (%)	Low risk (%)
1994–1999	1.20 ± 0.42	0	2.40 ± 0.70	1 (10%)	10 (100%)	0	0
2000–2005	1.69 ± 0.87	8 (34.8%)	4.09 ± 1.90	7 (30.4)	19 (82.6%)	4 (17.4%)	0
2006–2011	2.48 ± 1.38	19 (54.3%)	5.54 ± 2.57	19 (54.3%)	20 (57.1%)	10 (28.6%)	5 (14.3%)
P-value	0.004	0.005†	<0.001	0.022†		0.033†	
Total	2.03 ± 1.22	27 (39.7%)	4.59 ± 2.43	27 (39.7%)	49 (72.1%)	14 (20.6%)	5 (7.4%)

Kruskal–Wallis test, † $\chi^2$ -test.**Table 3** Quality assessment of RCT according to publication year (before and after 2001, IF introduction year)

Years	Jadad scale		van Tulder scale		Cochrane's assessment of risk bias		
	Score	High quality (%)	Score	High quality (%)	High risk (%)	Moderate risk (%)	Low risk (%)
1994–2000 ( <i>n</i> = 13)	1.31 ± 0.48	1/13 (7.7)	2.27 ± 1.01	2/13 (15.4)	13 (100%)	0	0
2001–2011 ( <i>n</i> = 55)	2.20 ± 1.28	26/55 (47.3)	5.02 ± 2.47	25/55 (45.5)	36 (63.6%)	14 (25.5%)	14 (25.5%)
P-value	0.013	0.007†	<0.001	0.043†		0.044†	

Kruskal–Wallis test, † $\chi^2$ -test.**Table 4** Characteristics of RCT according to subjects

Subjects, <i>n</i>	Jadad scale		van Tulder scale		Cochrane's assessment of risk bias		
	Score	High quality (%)	Score	High quality (%)	High risk (%)	Moderate risk (%)	Low risk (%)
Voiding dysfunction, 22	2.68 ± 1.32	12 (54.5)	5.77 ± 2.60	14 (63.6)	12 (54.5)	8 (36.4)	2 (9.1)
Oncology, 18	1.28 ± 0.57	3 (16.7)	2.94 ± 1.00	2 (11.1)	18 (100)	0	0
Stone/endourology/laparoscopic, 10	1.90 ± 1.52	3 (30.0)	4.30 ± 2.06	2 (20)	8 (80)	1 (10)	1 (10)
Infertility/andrology, 7	2.57 ± 1.27	4 (57.1)	5.86 ± 3.76	4 (57.1)	3 (42.9)	3 (42.9)	1 (14.3)
Infection, 3	1.33 ± 0.58	0	4.33 ± 1.52	1 (33.3)	2 (66.7)	0	1 (33.3)
Others, 8	1.87 ± 0.64	5 (62.5)	4.37 ± 1.50	4 (50.0)	6 (75.0)	2 (25.0)	0
P-value	0.004	0.051†	0.007	0.014†		0.048†	

Kruskal–Wallis test, † $\chi^2$ -test.

## Analysis of factors related to the quality of the articles

When the Jadad and van Tulder scales were used to assess quality, there were statistically significant differences in the mean scores observed between studies that consisted of drug treatment and non-drug treatment studies (Jadad  $2.29 \pm 1.34$ ,  $1.40 \pm 0.50$ ,  $P < 0.001$ ; van Tulder  $5.06 \pm 2.64$ ,  $3.45 \pm 1.23$ ,  $P = 0.001$ ). The ratios of high-quality articles among drug treatment RCT and non-drug articles were 19/48 (39.6%) and 8/20 (40%), respectively, when the Jadad scale was used, and 22/48 (45.8%) and 5/20 (25%), respectively, when the van Tulder scale was used. This indicates a slightly higher ratio of high-quality articles among drug treatment studies, but this difference was not statistically significant. When CCRBT was used, significantly more RCT that used drug treatments were assessed to be at low risk of bias ( $P = 0.024$ ). When the

Jadad and van Tulder scales were used to assess quality, the mean scores of studies with funding and IRB reviews were higher than studies without these factors. Furthermore, when CCRBT was used, more RCT with funding and IRB review were assessed to be at low risk of bias than the studies without these factors. However, only with the Jadad scale assessment was the ratio of high-quality articles not significantly different between studies with funding and those without. The mean scores of RCT that were carried out as multicenter studies were higher than non-multicenter studies on the Jadad and van Tulder scales, and the ratio of high-quality articles carried out as multicenter studies was higher only with the van Tulder assessment. However, when CCRBT was used to assess and compare whether the studies were carried out at multiple centers, the difference in the number of articles that were at low risk of bias was not significant (Table 5).

**Table 5** Factors associated with quality of RCT

Factors	No. RCT (%)	Jadad scale		van Tulder scale		Cochrane's assessment of risk bias		
		Score	High quality (%)	Score	High quality (%)	High risk (%)	Moderate risk (%)	Low risk (%)
Intervention type								
Drug	48 (70.59)	2.29 ± 1.34	19 (39.6)	5.06 ± 2.64	22 (45.8)	30 (62.5)	13 (27.1)	5 (10.4)
Non-drug	20 (29.41)	1.40 ± 0.50	8 (40.0)	3.45 ± 1.23	5 (25.0)	19 (95.0)	1 (1.5)	0
P-value		<0.001	0.974†	0.001	0.110†		0.024†	
Funding source								
Yes	15 (22.06)	2.93 ± 1.53	9 (60.00)	7.00 ± 2.95	11 (73.3)	6 (40.00)	5 (33.3)	4 (26.7)
No	53 (77.94)	1.77 ± 0.99	18 (34.0)	3.90 ± 1.76	16 (30.2)	43 (81.1)	9 (17.0)	1 (1.9)
P-value		0.013	0.069†	0.001	0.003†	0.001†		
Reviewed by IRB								
Yes	38 (55.88)	2.53 ± 1.33	20 (52.6)	5.79 ± 2.49	23 (60.5)	22 (57.9)	12 (31.6)	4 (10.5)
No	30 (44.12)	1.40 ± 0.67	7 (23.3)	3.07 ± 1.17	4 (13.3)	27 (90.0)	2 (6.7)	5 (7.4)
P-value		<0.001	0.014†	<0.001	<0.001†		0.013†	
Multicenter study								
Yes	30	2.47 ± 1.33	15 (50.0)	5.57 ± 2.90	16 (53.3)	19 (63.3)	8 (26.7)	3 (10.0)
No	38	1.68 ± 1.02	12 (31.6)	3.81 ± 1.64	11 (28.9)	30 (78.9)	5 (15.8)	5 (7.4)
P-value		0.008	0.123†	0.002	0.041†		0.360†	

Student's t-test, † $\chi^2$ -test.

## Discussion

Since the first RCT was published in IJU in 1994, the total number of RCT published has gradually increased, but the ratio of RCT relative to all articles has not significantly increased. The number and percentage of high-quality RCT published in IJU has increased over time. However, some of the research papers did not clearly state the methods or carry out randomization properly. Furthermore, some studies inappropriately carried out blinding, concealment of allocation and intention to treat analyses.

There are various tools that can be used to evaluate the quality of RCT, including the Campell, Moher, Chalmers, Jadad, van Tulder, Newell's and Cochrane assessments. We used the widely-used Jadad scale, van Tulder scale and CCRBT methods that have been recommended by several meta-analysis groups to evaluate the quality of RCT. Recently, various studies have attempted to assess the quality of RCT using these tools. Lee *et al.* analyzed 28 (0.89%) RCT published in KJU from 1991 to 2010 by using the Jadad scale only. They reported that the quantity and mean Jadad scale score of RCT has gradually increased over time. Lee *et al.* also suggested two items (descriptions of double blinding and dropouts) in which RCT published in KJU needed improvement.<sup>13</sup> Similarly, in the present study, the mean Jadad scale score for RCT published in IJU increased according to time. Chung *et al.* analyzed RCT published in KJFM from 1980 to 2005,<sup>14</sup> and found that the mean scores of RCT also increased over time. Although Chung *et al.* only used the Jadad scale to assess the quality of RCT in KJFM, they also analyzed allocation concealment and found that there were just two (8.7%) articles where proper allocation concealment was implemented. In the

present study, there were just six (8.8%) RCT in IJU that described properly-designed allocation concealment.

As the importance of EBM has been gradually highlighted, the high level of evidence providing RCT is estimated to be increasing.<sup>15</sup> Although these previous studies did not compare the quality of RCT over time, the present study showed that the scale scores and the ratios of high-quality articles in IJU increased in quality over time when assessed by the Jadad scale, the van Tulder scale and CCRBT.

When the quality of articles was compared in single years rather than 6-year units, the improvements in quality were observed with the van Tulder scale, but not the Jadad scale (Fig. 1). These differences in RCT quality are a result of the various differences between the assessment methods and standards. The Jadad scale scoring is too simple and places too much emphasis on blinding. The van Tulder scale provides more scrutiny than the Jadad scale, but all 11 items in the van Tulder scale are weighted equally. When CCRBT is used to assess quality, allocation concealment, sequence generation and blinding items must all be satisfied for an article to be assessed with a lower risk of bias. Even if proper randomization is carried out, there remains a probability of bias for patients, researchers and related researchers if allocation is not securely concealed.<sup>16</sup> Schulz *et al.* reported that the effects of interventions in RCT could be overstated up to 40% when allocation was not properly concealed.<sup>17</sup> The first article in IJU that showed proper allocation concealment was published in 2004. Among the 11 RCT published after 2010, just two (18.1%) showed proper allocation concealment. Because of an increase in appropriate implementation of double blinding, randomization and dropouts over time, the ratios of high-quality articles increased with the Jadad and van Tulder scales. However, the continuous



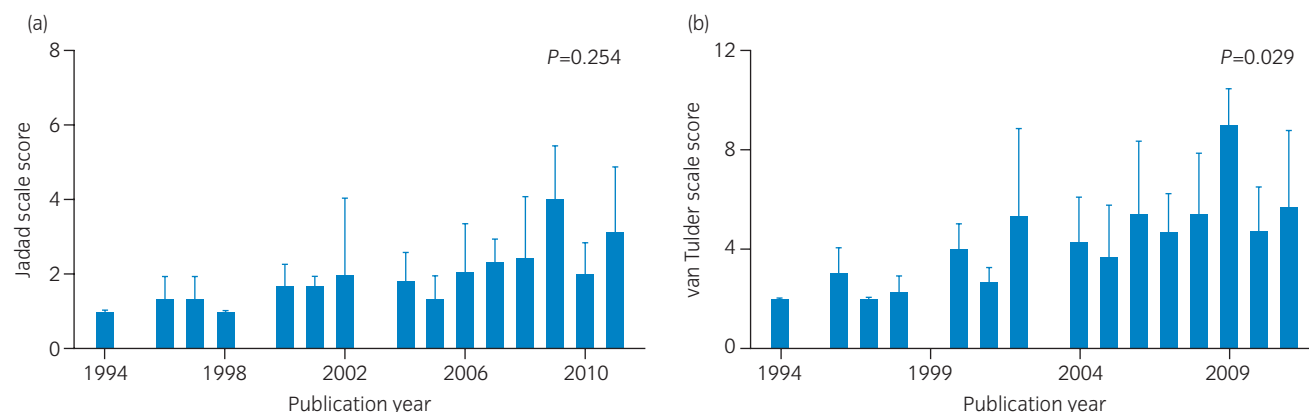


Fig. 1 Quality assessment of RCT according to publication year. One-way ANOVA analysis of (a) Jadad scale score and (b) van Tulder scale score.

implementations of improper allocation concealment methods resulted in a small number of articles with a low risk of bias when assessed by CCRBT.

Furthermore, the present study confirmed that RCT published in IJU that underwent IRB review had higher quality ratios relative to RCT that did not. To date, no other study has been published that investigated the relationship between the presence of IRB review and article quality. Proper arrangement of the research methods and implementation of IRB review in the planning process might increase the validity of a study. A study that did not use drugs as a treatment modality had difficulty in keeping both patients and investigators blind. The present study found that blinding was not carried out properly in articles that did not use drug treatments and, as a result, only a small number of high-quality articles were observed when compared with RCT that did use drug interventions. The present study also found that published RCT with funding were assessed to be higher in quality than studies without funding. There is controversy over assessing the quality of articles based on funding resources, because an insufficient number of assessments have been carried out. According to the Lee *et al.* study that analyzed the quality of RCT in KJU, the studies that received funding resulted in a higher ratio of high-quality articles, because their investigation schemes were better prearranged.<sup>13</sup> However, Clifford *et al.* analyzed 100 RCT published in five leading general medical journals, and reported that there was no correlation between funding resources and the quality of articles.<sup>18</sup> In the present study, articles with funding resources were assessed to have higher scores with Jadad and van Tulder scale assessment, and higher ratios of high-quality articles with the van Tulder scale assessments than articles without funding resources. Furthermore, the ratio of articles with funding had a significantly lower risk of bias with the CCRBT assessment.

RCT that were carried out in multiple centers were evaluated to be high quality. Prearrangement of research schema to carry out uniform data gathering and analyzing in several institutions could be beneficial for improving the quality of

articles. However, as only a limited number of multicenter RCT were examined in the present study, this conclusion is not valid until further assessments with supplementary RCT can be carried out.

The numbers of RCT published in IJU on voiding dysfunction, oncology and stone disease/endourology/laparoscopy were 22, 18 and 10, respectively. These articles accounted for 73.5% of all RCT. Voiding dysfunction was the subject that had the largest number of articles (22 RCT; 32.3%) and the highest ratio of high quality articles (Jadad: 54.5%; van Tulder: 63.6%). Among the studies on voiding dysfunction, articles that used drug interventions, which for this subject it is relatively easy to carry out both randomization and double blinding, and articles that underwent IRB reviews were numerous. For these reasons, the quality of articles about voiding dysfunction was increased.

In the field of oncology, it is difficult to carry out RCT for studies of novel antitumor agents or treatment methods; furthermore, these studies often pose ethical problems. We believe this is why such studies are rarely published in most journals. Therefore, only RCT that have minimal expected complications or ethical issues are likely to be published. Such a trend was seen in the RCT published in IJU. Of the 18 oncology studies published in this journal, eight were on pre- or postoperative bladder instillation after transurethral resection of bladder tumor, nine were on androgen deprivation therapy before and after radical prostatectomy, and one was on adjuvant therapy with UFT (a 1:4 mixture of tegafur and uracil) after radical nephrectomy. Doubling blinding could not be carried out for most of these studies. Because of this limitation, studies that make blinding impossible are categorized as high-quality articles in the Jadad scale if they score two or more. However, the van Tulder scale and the CCRBT do not consider such exceptions. We look forward to a consensus for new assessment tools that correct for these limitations in future discussions.

IF is a measure of the relative importance of a journal, and journals with a higher IF tend to publish articles of higher

quality. Barbui *et al.* used RCT comparing two groups of patients with depression who were either being treated with fluoxetine or with another pharmacological agent to compare the quality of RCT according to the IF of the journal.<sup>19</sup> The authors concluded that the IF was not associated with the quality of the RCT. However, as that is a study of a small number of RCT for a specific disease, it cannot represent the trend of all RCT. Journals with higher IF tend to prefer to publish higher-quality articles, and require well-designed high-quality RCT for publication. With the increasing importance of EBM in the literature, the demand for high-quality RCT is growing. We believe that the IF of a journal affects the quality of the RCT it publishes. We look forward to future studies in this area being published. The present study showed that the quality of published RCT increased after the introduction of IF in 2001. This might be a result of the overall improvement in the quality of articles over time; however, we believe it is also a result of attempts by IJU to maintain and improve the quality of the work published after the introduction of IF; and also the introduction of the new Good Clinical Practice in 1998 could improve the quality of physicians initiating clinical trials in Japan. Such changes are likely to be another reason for the gradual increase in the quality of RCT.

The present study is not without limitations. As the study was manually researched and evaluated, subjective opinions might have been involved in the assessment process. To compensate for this limitation, two designated medical practitioners participated in sampling individually, and two specialists carried out the quality assessments independently. The authors of the papers need to be blinded, as any affiliation that the evaluator has with the authors might affect their quality assessment. The purpose of this process was to adjust the significant differences in the results and to secure the objectivity and reliability of the study. The other limitation is that there is no consensus on which quality assessment tools are the most accurate. However, in the present study, we made an effort to overcome this limitation by using three distinct and representative quality assessment tools. The present study was able to show both qualitative and quantitative changes in RCT published in IJU from 1994 to 2011 by carrying out quality assessments. The objective of the study was to suggest research directions and to facilitate improvements in the quality of articles published in IJU. In the present study, the quality of the papers was compared using criteria such as whether funding was provided and when the paper was published. The amount of funding, length of the study period and the year of protocol development are likely to be associated with the quality of the RCT. However, those factors were not stated, meaning these factors could not be analyzed.

The total number of original articles and RCT has increased over time since the IJU was first issued in 1994. Although there have been no significant changes in the

ratios of RCT relative to the total number of articles published in IJU, the quality of RCT has increased. A sustained effort to enhance the research quality and increase the number of RCT will contribute to the advancement of IJU.

## Conflict of interest

None declared.

## References

- 1 Egger M, Jüni P, Bartlett C; CONSORT Group (Consolidated Standards of Reporting of Trials). Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001; **285**: 1996–9.
- 2 Begg C, Cho M, Eastwood S *et al.* Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996; **276**: 637–9.
- 3 Plint AC, Moher D, Morrison A *et al.* Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med. J. Aust.* 2006; **185**: 263–7. Epub 2006/09/05.
- 4 Clark HD, Wells GA, Huet C *et al.* Assessing the quality of randomized trials: reliability of the Jadad scale. *Control. Clin. Trials* 1999; **20**: 448–52. Epub 1999/09/30.
- 5 van Tulder M, Furlan A, Bombardier C, Bouter L; Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine (Phila Pa 1976)* 2003; **28**: 1290–9. Epub 2003/06/18.
- 6 Higgins J, Green S, Cochrane C, Cochrane Reviewers H. *Cochrane Handbook for Systematic Reviews of Interventions*, Version 5.0.2. [S].l. Cochrane Collaboration, John Wiley & Sons, Hoboken, 2008.
- 7 Jadad AR, Moore RA, Carroll D *et al.* Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control. Clin. Trials* 1996; **17**: 1–12. Epub 1996/02/01.
- 8 Graham N, Haines T, Goldsmith CH *et al.* Reliability of 3 assessment tools used to evaluate randomized controlled trials for treatment of neck pain. *Spine (Phila Pa 1976)* 2012; **37**: 515–22. Epub 2011/06/16.
- 9 Spooren AI, Janssen-Potten YJ, Kerckhofs E, Seelen HA. Outcome of motor training programmes on arm and hand functioning in patients with cervical spinal cord injury according to different levels of the ICF: a systematic review. *J. Rehabil. Med.* 2009; **41**: 497–505. Epub 2009/06/23.
- 10 Langhorst J, Musial F, Klose P, Hauser W. Efficacy of hydrotherapy in fibromyalgia syndrome – a meta-analysis of randomized controlled clinical trials. *Rheumatology (Oxford)* 2009; **48**: 1155–9. Epub 2009/07/18.
- 11 Zech A, Hubscher M, Vogt L, Banzer W, Hansel F, Pfeifer K. Balance training for neuromuscular control and performance enhancement: a systematic review. *J. Athl. Train.* 2010; **45**: 392–403. Epub 2010/07/14.
- 12 Green S, Higgins JPT, Alderson P, Clarke M, Mulrow C, Oxman A. *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane Collaboration, John Wiley & Sons, Hoboken, 2008.
- 13 Lee JY, Chung JH, Kang DH *et al.* Quality assessment of randomized controlled trials published in the Korean journal of urology over the past 20 years. *Korean J. Urol.* 2011; **52**: 642–6.
- 14 Chung W, Lee KW, Hwang IH, Lee DH, Kim SY. Quality assessment of randomized controlled trials in the Journal of the Korean Academy of Family Medicine. *Korean J. Fam. Med.* 2009; **30**: 626–31.
- 15 Keech AC, Pike R, Granger RE, Gebbski VJ. Interpreting the results of a clinical trial. *Med. J. Aust.* 2007; **186**: 318–19. Epub 2007/03/21.
- 16 Viera AJ, Bangdiwala SI. Eliminating bias in randomized controlled trials: importance of allocation concealment and masking. *Fam. Med.* 2007; **39**: 132–7. Epub 2007/02/03.
- 17 Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet* 2002; **359**: 614–18. Epub 2002/02/28.
- 18 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. Epub 2002/09/06.
- 19 Barbui C, Cipriani A, Malvini L, Tansella M. Validity of the impact factor of journals as a measure of randomized controlled trial quality. *J. Clin. Psychiatry* 2006; **67**: 37–40. Epub 2006/01/24.