Review Article

Randomized Controlled Trials of Acupuncture (1997–2007): An Assessment of Reporting Quality with a CONSORT- and STRICTA-Based Instrument

Richard Hammerschlag,1 Ryan Milley,1 Agatha Colbert,2 Jeffrey Weih,3 Beth Yohalem-Ilsley,1 Scott Mist,4 and Mikel Aickin5

1 Research Department, Oregon College of Oriental Medicine, Portland, OR 97216-2859, USA
2 Helfgott Research Institute, National College of Natural Medicine, Portland, OR 97201-4848, USA
3 Department of Physical Medicine & Rehabilitation, Kaiser Permanente Northwest, Portland, OR 97227-1110, USA
4 Fibromyalgia Research Unit, Oregon Health Science University, Portland, OR 97239-3098, USA
5 Program in Integrative Medicine and Department of Family & Community Medicine, College of Medicine, University of Arizona, Tucson, AZ 85724-5052, USA

Correspondence should be addressed to Richard Hammerschlag, rhammerschlag@ocom.edu

Received 4 March 2010; Accepted 1 July 2010

Copyright © 2011 Richard Hammerschlag et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The present study describes the development of a comprehensive quality of reporting assessment tool and its application to acupuncture RCTs from 1997–2007. This Oregon CONSORT STRICTA Instrument (OCSI) is based on the revised CONSORT guidelines as modified by the STRICTA recommendations for acupuncture trials. Each of the resulting 27 OCSI items were applied to English language prospective RCTs that compared acupuncture, using manual and/or electro-stimulation, to no treatment, a sham procedure, or usual biomedical care. The 333 RCTs that met inclusion criteria were dispersed among 27 countries and 141 journals. Mean quality of reporting score for all articles was 63.0% (SD 16.5). Mean OCSI scores revealed a 30.9% improvement over the ten-year period ($P<.001$). Our findings suggest that to enhance quality of reporting, authors should better attend to seven specific OCSI items in three categories: practitioner training, adverse events, and aspects of randomization and blinding ($n = 5$). The broad diversity in geographical origin, publication site and quality of reporting, viewed in light of the considerable room for improvement in mean OCSI scores, emphasizes the importance of making STRICTA as well as CONSORT more widely known to journals and to the acupuncture research community.

1. Introduction

Systematic reviews represent a retrospective, criteria-based approach for summarizing research findings [1–3]. By applying predetermined standards to identify the trials to be reviewed, and uniform criteria to evaluate the selected trials, bias in the quality assessment process is minimized. With the evidence based perspective being increasingly applied to complementary and alternative medicine (CAM) in the past decade, a wide variety of condition-focused systematic reviews have evaluated randomized controlled trials (RCTs) of acupuncture. MEDLINE alone lists over 170 of such reviews through 2009, many of which have, in turn, been summarized and analyzed [4–10]. As noted in these overviews, systematic reviews of acupuncture have employed a heterogeneous group of quality assessment instruments, that vary from the 5-item Jadad scale [11] and a modified 6-item Jadad scale [12–14], to the Cochrane Collaboration guidelines [15] and a range of broader scales containing up to 27 items [16].

In the present paper, we describe the development and application of a comprehensive quality of reporting instrument for rating RCTs of acupuncture, based on the revised CONSORT guidelines for RCTs [17] as modified by the STRICTA recommendations for acupuncture trials [18]. The Consolidated Standards of Reporting Trials
(CONSORT) statement was created as a set of guidelines for use by journal editors, reviewers, and authors to increase the likelihood that RCTs submitted for publication would meet uniform standards for reporting [19]. The Standards for Reporting Interventions in Controlled Trials of Acupuncture (STRICTA) were crafted to modify a single item on the 22-item CONSORT list, referring to description of interventions. This item (CONSORT no. 4), calling for “precise details of the interventions for each group and how they were actually administered,” was considered too generic to be of value for improving reporting of acupuncture trials and was expanded to a 6-item list, with each item broken out into subitems, specifying the details of the acupuncture protocol to be reported [18]. Thus, STRICTA was created to complement, not substitute for, CONSORT.

It is important to recognize that the CONSORT statement is described as “…a tool to improve quality of reporting of RCTs… (but not) …as a formal quality assessment instrument.” [17, 20]. Despite this disclaimer, over 30 systematic reviews (predominantly focused on biomedical RCTs) have converted the CONSORT guidelines to a variety of scoring systems for evaluating quality of reporting. While several of these adaptations involved only a limited set of “essential” CONSORT items [21–24], many of the reviews devised a scoring system that utilized the full complement of CONSORT items [25–27], with some also breaking out selected items to create larger lists [28–32]. One review, published in Chinese, is the first to create a combined CONSORT and STRICTA quantitative assessment tool, applying it to the worldwide literature on RCTs of acupuncture for obesity [33].

A recent review sought to assess and compare the impact of CONSORT and STRICTA on the reporting quality in acupuncture trials during three time periods 1994-1995, 1999-2000, and 2004-2005 [34]. The review concluded that the reporting of selected CONSORT items has improved over time whereas no significant improvements were observed in STRICTA items. The authors also state that further exploration of the adherence to CONSORT and STRICTA within acupuncture RCTs is warranted, due to the limited number of studies sampled.

The present study describes the independent creation of a combined CONSORT- and STRICTA-based quality of reporting assessment tool, the Oregon CONSORT STRICTA Instrument (OCSI), and its application to acupuncture RCTs, across all conditions, published in the ten-year period following the October, 1997 NIH Consensus Development Conference on Acupuncture [35]. Our aims were to (1) develop a comprehensive quality of reporting assessment instrument based on two existing guidelines for clinical trial reporting, (2) examine mean scores per question, to inform the acupuncture research community as to which research design items are most often poorly reported or omitted, and (3) examine the overall scores per trial, to provide an indication of whether the quality of reporting in acupuncture RCTs has been improving over time. The development and application of OCSI have been presented in preliminary form [36].

2. Methods

2.1. Development of the Oregon CONSORT STRICTA Instrument (OCSI). In creating the OCSI quality of reporting assessment tool, we followed the STRICTA recommendations of substituting 6 items, relevant to acupuncture and control group interventions, for item no. 4 of the 22 CONSORT items. Each of the resulting 27-item combined CONSORT and STRICTA guidelines was then converted to a question, retaining the CONSORT item sequence and wherever possible preserving the original wording for each item (Table 1). When clarification was necessary, we modified sentence structure but strove to remain as close as possible to the original wording. Where more significant changes were considered helpful, they were added to a separate list of modifications for future consideration.

A more significant problem in converting evaluative guidelines to questions was that most of the CONSORT and STRICTA items contained several embedded subitems that, while related, required modification to separate, subquestions. For example, CONSORT item no. 11, concerned with reporting of blinding, states:

“Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. If done, how the success of blinding was evaluated.” [17]

This compound item was converted to a multipart OCSI question, asking:

“Is it stated whether (a) participants, (b) those administering the interventions, and (c) those assessing the outcomes were blinded to group assignment; and (d) was the success of blinding evaluated?”

Each of the 27 OCSI questions was constructed to be scored yes = 2, partial = 1, no = 0, or Not Applicable = N/A, based on the composite scoring of its sub-questions. A question was scored partial if its sub-questions received a mix of ratings. If one or more sub-questions were scored N/A, the score for the question was based on the scores of the remaining sub-questions. N/A also could be assigned to an individual question; for example, OCSI question no. 7 regarding reporting of cointerventions was scored N/A if no adjunctive treatment (e.g., herbal medicine or moxibustion) was provided to the acupuncture group.

2.2. Article Selection Criteria. RCTs included for assessment met the following criteria: (1) publication date from November 1997 through October 2007; (2) prospective, randomized controlled trial; (3) human subjects; (4) English language; (5) full publication; (6) treatment with filiform acupuncture needles using manual and/or electrostimulation; (7) comparator/control group consisted of no treatment, a sham procedure, or usual biomedical care. Usual biomedical care was defined a priori as interventions usual and customary to biomedical conditions. These interventions were categorized...
<table>
<thead>
<tr>
<th>Item number</th>
<th>Paper section</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Abstract</td>
<td>Is there an explicit statement that patients were randomly assigned to interventions?</td>
</tr>
<tr>
<td>2</td>
<td>Introduction/Background</td>
<td>(a) Is scientific background provided and (b) is the rationale explained?</td>
</tr>
<tr>
<td>3</td>
<td>Methods</td>
<td>(a) Are the eligibility criteria (inclusion and exclusion criteria) stated and (b) are the setting(s) and location(s) where the data was collected described?</td>
</tr>
<tr>
<td>4</td>
<td>Methods</td>
<td>(a) Is the style of acupuncture stated? (b) Is the rationale presented for the selection of acupuncture points? (c) Was the rationale justified?</td>
</tr>
<tr>
<td>5</td>
<td>Methods</td>
<td>Are the following parameters of needling presented? (a) Points used (uni/bilateral) (b) Number of needles inserted (c) Depth(s) of insertion (d) Response elicited (e.g., de qi) (e) Needle stimulation (manual or electrical) (f) Needle retention time (g) Needle type (Material and/or manufacturer, gauge, and length)</td>
</tr>
<tr>
<td>6</td>
<td>Methods</td>
<td>Are the (a) number and (b) frequency of treatments stated?</td>
</tr>
<tr>
<td>7</td>
<td>Methods</td>
<td>Are details of the acupuncture group cointervention(s) presented? (e.g., moxa, cupping, life-style advice, plum-blossom needling, Chinese herbs)</td>
</tr>
<tr>
<td>8</td>
<td>Methods</td>
<td>Are descriptions provided of the (a) duration of practitioner training, (b) length of clinical experience, and (c) expertise in specific condition?</td>
</tr>
<tr>
<td>9</td>
<td>Methods</td>
<td>(a) Is the intended effect of the control or comparison intervention presented? (b) Were the specific explanations given to patients of the treatment and control interventions presented? (c) Are details for the control or comparison intervention presented? (d) Are sources provided, that justify the choice of the control or comparison intervention?</td>
</tr>
<tr>
<td>10</td>
<td>Methods</td>
<td>Are there statements of (a) specific objectives and (b) hypotheses to be tested?</td>
</tr>
<tr>
<td>11</td>
<td>Methods</td>
<td>(a) Are primary and (if applicable) secondary outcome measures clearly defined? (b) Are there statements (when applicable), regarding any methods used to enhance the quality of measurements, for example, multiple observers or training of assessors?</td>
</tr>
<tr>
<td>12</td>
<td>Methods</td>
<td>(a) Is there a statement regarding how the sample size was determined, and (b) if applicable, an explanation of any interim analyses and stopping rules?</td>
</tr>
<tr>
<td>13</td>
<td>Methods</td>
<td>(a) Is the method presented that was used to generate the random allocation sequence, and (b) if applicable, details of any restriction (e.g., blocking, stratification)?</td>
</tr>
<tr>
<td>14</td>
<td>Methods</td>
<td>(a) Is the method presented that was used to implement the random allocation sequence, (b) with clarification as to whether the sequence was concealed until interventions were assigned?</td>
</tr>
<tr>
<td>15</td>
<td>Methods</td>
<td>Are there statements as to (a) who generated the allocation sequence, (b) who enrolled participants, and (c) who assigned participants to their groups?</td>
</tr>
<tr>
<td>16</td>
<td>Methods</td>
<td>Is it stated whether or not (a) participants, (b) those administering the interventions, and (c) those assessing the outcomes were blinded? and (d) was the success of participant blinding evaluated?</td>
</tr>
<tr>
<td>17</td>
<td>Methods</td>
<td>(a) Were the statistical methods stated that were used to compare groups for primary outcomes? (b) Were the statistical methods stated that were used for additional analyses such as subgroup or adjusted analyses?</td>
</tr>
<tr>
<td>18</td>
<td>Results</td>
<td>(a) Is the flow of participants through each stage quantitatively described, and (b) if protocol deviations are reported, were reasons presented?</td>
</tr>
<tr>
<td>19</td>
<td>Results</td>
<td>(a) Are dates provided that define the period of recruitment? (b) Is the length of followup (on-treatment and posttreatment) reported?</td>
</tr>
<tr>
<td>20</td>
<td>Results</td>
<td>(a) Are baseline demographics and (b) clinical characteristics presented for each group?</td>
</tr>
<tr>
<td>21</td>
<td>Results</td>
<td>(a) Is the number of participants in each group included in each analysis? (b) Was the “intention to treat” analysis presented? (c) When feasible, are the results stated in absolute numbers (e.g., 10 of 20, not just 50%)?</td>
</tr>
</tbody>
</table>
as educational, behavioral, physical, or pharmaceutical. RCTs were excluded if acupuncture points were stimulated by means other than filiform needles, for example, ear tacks, intradermal needles, TENS, or laser.

Databases that were searched to identify articles included MEDLINE, the Cochrane Central Register of Controlled Trials, Alt HealthWatch, AMED, University of Maryland CAMPAIN, and the Oregon College of Oriental Medicine library database, which includes 16 non-MEDLINE journals of acupuncture and Oriental medicine. In addition, hand searches were performed of the reference lists from the WHO report on controlled clinical trials of acupuncture [37] and from 71 systematic reviews of acupuncture published between October 1997 and October 2007.

2.3. Application of OCSI. Seven individuals with experience in acupuncture research formed the OCSI group. Two members served as alternates, ensuring a group of 5 OCSI-raters over the course of the project. The group comprised a diverse range of backgrounds, all with experience in assessing RCTs of acupuncture, including three licensed acupuncturists, two nonpractitioners, a medical acupuncturist, and an acupuncture student. As a means of eliminating systematic bias, it was initially suggested to blind reviewers of RCTs to article information such as journal, author, and publication year [11]. An attempt to validate this recommendation was unsuccessful [38], and a recent assessment in acupuncture trials demonstrated limited difference in outcomes [34]. Based on these results, the OCSI Group was not blinded to article information.

From the RCTs that met the selection criteria, eight articles were initially chosen at random for scoring by the OCSI group (each of five raters) to assess face validity of the OCSI questions. Several group meetings were held to compare raters’ scores per question on each of the eight articles and to seek consensus on scoring. The latter process included development of an OCSI manual that outlines criteria for scoring each question as yes, partial, no, or N/A (Appendix A).

Following the consensus process, the remaining articles were randomly distributed among five raters for individual scoring. Each article’s score was divided by the total possible score (2 \times [27 – n questions scored N/A]) and converted to percent. Scores for each OCSI question were entered into a database from which total OCSI scores per article were calculated. A second database contained extracted demographics from each article (e.g., country of origin, journal, year, condition). Results for the first eight articles were group consensus scores; results for the remaining articles were from single raters.

2.4. Statistics. For descriptive purposes, we computed and displayed averages of the individual question scores. For select questions, regression analysis was performed, and binomial 95% confidence intervals were computed for change of reporting over time. The distribution of all RCTs across OCSI score by decile is presented as a histogram.

Interrater agreement was assessed at a point in the study when low, middle, and high scoring articles could be discerned. Nine RCTs were randomly selected, three each from each of these categories, in order to estimate agreement across the spectrum of quality. Conventional intraclass coefficients (ICCs) were computed from results with five reviewers each rating all nine articles.

Since China contributed the largest number of articles and had the lowest mean OCSI score by country, we plotted the mean OCSI scores by year for all countries, for China alone, and for all countries other than China, and fitted linear trends. Regression analysis was initially performed for all countries combined. A second regression analysis included a binary variable for China or all others as a covariate.
Table 2: Distribution of OCSI scores by country.

<table>
<thead>
<tr>
<th>Country</th>
<th>Articles (n)</th>
<th>Mean Score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>78</td>
<td>45.2</td>
</tr>
<tr>
<td>USA</td>
<td>58</td>
<td>69.7</td>
</tr>
<tr>
<td>Germany</td>
<td>45</td>
<td>73.2</td>
</tr>
<tr>
<td>Sweden</td>
<td>33</td>
<td>64.4</td>
</tr>
<tr>
<td>UK</td>
<td>23</td>
<td>72.5</td>
</tr>
<tr>
<td>Norway</td>
<td>14</td>
<td>68.7</td>
</tr>
<tr>
<td>Italy</td>
<td>11</td>
<td>59.6</td>
</tr>
<tr>
<td>Taiwan</td>
<td>11</td>
<td>58.8</td>
</tr>
<tr>
<td>Australia</td>
<td>10</td>
<td>72.3</td>
</tr>
<tr>
<td>Austria</td>
<td>7</td>
<td>61.6</td>
</tr>
<tr>
<td>Japan</td>
<td>7</td>
<td>62.8</td>
</tr>
<tr>
<td>Spain</td>
<td>4</td>
<td>85.4</td>
</tr>
<tr>
<td>Brazil</td>
<td>3</td>
<td>71.2</td>
</tr>
<tr>
<td>Croatia</td>
<td>3</td>
<td>46.7</td>
</tr>
<tr>
<td>Denmark</td>
<td>3</td>
<td>63.7</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>3</td>
<td>85.0</td>
</tr>
<tr>
<td>Iran</td>
<td>3</td>
<td>61.4</td>
</tr>
<tr>
<td>Israel</td>
<td>3</td>
<td>59.3</td>
</tr>
<tr>
<td>South Korea</td>
<td>3</td>
<td>60.9</td>
</tr>
<tr>
<td>Canada</td>
<td>2</td>
<td>78.5</td>
</tr>
<tr>
<td>Switzerland</td>
<td>2</td>
<td>75.3</td>
</tr>
<tr>
<td>Turkey</td>
<td>2</td>
<td>62.7</td>
</tr>
<tr>
<td>France</td>
<td>1</td>
<td>80.0</td>
</tr>
<tr>
<td>Greece</td>
<td>1</td>
<td>48.0</td>
</tr>
<tr>
<td>Ireland</td>
<td>1</td>
<td>65.0</td>
</tr>
<tr>
<td>Singapore</td>
<td>1</td>
<td>72.0</td>
</tr>
<tr>
<td>Thailand</td>
<td>1</td>
<td>62.0</td>
</tr>
<tr>
<td>Total</td>
<td>333</td>
<td>63.0</td>
</tr>
</tbody>
</table>

3.2. OCSI Manual and Interrater Agreement. The initial face validity and consensus-building exercise led to clarification of OCSI items and drafting of a manual to inform scoring decisions. In most cases, the “Example and Explanation” section of the CONSORT document [17] and the similar section from STRICTA [18] were sufficient to facilitate its development. The manual proved especially helpful to set parameters for deciding between a score of partial versus yes, for example, item no. 22(c): “Presentation of \( P \) values alone will score partial; for full credit, authors must state precision as a confidence interval (CI).” The OCSI manual is presented as Appendix A.

3.3. OCSI Scores per Question and per Article. OCSI scores per individual question across all trials are presented in Figure 1. Since a rating of partial is scored as a 1, it is of interest that 7 of the 27 questions showed mean scores <1.0 (ratings between no and partial). Of these seven, only one was of STRICTA origin, question 8, asking for information on practitioner training, experience, and expertise. The other low-scoring questions are those that ask about reporting of sample size calculation (no. 12), randomization generation (no. 13), randomization implementation (no. 14), person responsible for randomization (no. 15), blinding (no. 16), and adverse events (no. 24). Of these poorly reported questions, 4 of 7 demonstrate trends of improvement over time, 3 of which reached statistical significance. These are questions: 12 (3.5% improvement \( P<.001; 95\% CI=0.000–0.071 \)); 13 (4.9% \( P<.001; 95\% CI=0.014–0.084 \)); 14 (6.2% \( P<.001; 95\% CI=0.028–0.094 \)); 24 (4.6% \( P<.016; 95\% CI=0.009–0.083 \)).

In contrast, 9 questions showed means ≥1.5 (of a possible 2.0) (Figures 1(a) and 1(b)). The highest scorers among these (in order of ranking) were questions asking for reporting of number and frequency of treatments (no. 6), primary and secondary outcomes (no. 11), scientific background, and rationale (no. 2), statistical methods (no. 17) and randomization included in title or abstract (no. 1).

Distribution of OCSI scores per RCT is presented in Figure 2. The mean percent score for all articles was 63.0 (SD 16.5). The numbers of trials with OCSI scores higher than arbitrary values of 60, 70, and 80% were 197 (59.2%), 133 (39.9%), and 62 (18.6%), respectively. Mean score of RCTs from CAM journals (51.3% SD ±15.1) was significantly lower than the mean from biomedical journals (69.9% ± 13.1) (\( P<.001 \); 95% CI = 15.4–21.8%). Of the five countries publishing the greatest number of RCTs (China, USA, Germany, Sweden, and UK), the mean OCSI score of articles from China (45.2% ± 14%) was significantly lower (\( P<.001 \); 95% CI = 21.0–28.2%) than mean scores of articles from the other four countries (69.8% ± 2.7). The top five journals publishing trials that met our inclusion criteria are J Tradit Chin Med (n = 39), Internat J Clin Acupunct (n = 14), Pain (n = 13), J Altern Complement Med (n = 12), and

3. Results

3.1. Demographics of Included Trials. An initial search of the literature identified 410 RCTs that appeared to meet the inclusion criteria. Further review of the papers excluded 77 trials for reasons ranging from trials that were duplicate publications to trials that compared two forms of acupuncture. Inclusion criteria, including publication date from November 1997 through October 2007, were met by 333 acupuncture RCTs (Appendix B). These trials represented 27 countries with China accounting for the largest share (\( n = 78; 23.4\% \)). Trials from the top five countries combined, including China, United States, Germany, Sweden, and United Kingdom, totaled 237 (71.2%) (Table 2). Asian countries other than China accounted for 26 (6.9%) trials. No attempt was made to determine how many RCTs were excluded on the basis of non-English language publication.

The 333 RCTs appeared in 141 journals, and the large majority of articles were listed in MEDLINE (\( n = 302; 90.7\% \)). Of the total included trials, almost two-thirds (\( n = 209; 62.7\% \)) appeared in biomedical journals, with the remainder (\( n = 124 \)) in journals of complementary and alternative medicine.
Acupunct Med (n = 11). Of these five journals, The Journal of Traditional Chinese Medicine, which accounted for the largest percentage of publications (11.7%), demonstrated the lowest mean OCSI score for its acupuncture RCTs (39.6%).

Figure 3 shows mean OCSI and modeled scores over time for trials from all countries, from China alone, and from countries other than China. The mean OCSI scores from all countries demonstrate a significant improvement of 30.9% over the ten-year period (P < .001). Significant improvements are found in RCTs from China 37.3% (P < .005) and countries other than China 32.2% (P < .005) over time. The modeled mean 1998 OCSI score of RCTs from China (36.1%) is significantly lower (P < .01) than the corresponding score from countries other than China (59.6%). The improvement on OCSI scores from China (1.4% per year, P < .005) as compared to the improvement in all other countries (1.7% per year, P < .001) is not significantly different (P < .42).

4. Discussion

Creation of OCSI required two main steps: conversion of the combined guidelines into questions, including breakout of each multicomponent item into a nested set of sub-questions, and development of a manual that outlines criteria for scoring each question as yes, partial, no, or N/A. The former step resulted in a lengthier scoring instrument than we had initially envisioned (63 sub-questions grouped into 27 main questions). The latter, based in large part on the detailed rationales presented with CONSORT [17] and STRICTA, proved essential for applying OCSI and achieving interrater agreement.

The acupuncture RCTs identified for scoring by OCSI revealed a strikingly broad diversity in both geographical origin and publication site, a finding supported by the recent review by Prady et al. [34]. These demographics, viewed in light of the considerable room for improvement in mean percentage OCSI scores of acupuncture trials appearing in both CAM (51.3, ±15.1) and biomedical (69.9, ±13.1)
journals, emphasize the importance of making CONSORT and STRICTA more widely known and applied. This point is highlighted by a recent survey of high impact journals \((n = 165)\), of which only 62 (38%) mentioned the CONSORT statement in their online “Instructions to Authors” while 23 (14%) stated that a completed CONSORT statement was a condition of submission \([39]\). Lack of adequate reporting is not unique to English language publications. A recent review of adherence to CONSORT in 142 RCTs from five leading Chinese medical journals indicated an overall low quality of reporting \([40–42]\). The translations of both CONSORT and STRICTA into several Asian languages should help the dissemination effort \([43–46]\).

In regard to the seven lowest-scoring OCSI questions (those with mean score <1.0), several recommendations can be made to the acupuncture research community. The sole question from STRICTA within this group (question no. 8) is important since practitioner training and experience may significantly affect the outcome of a trial. The low reporting of this item is consistent with the recent assessment of STRICTA items \([34]\) and is also reflected in the low ranking of the utility of this question by authors of acupuncture trials \([47]\). The importance of reporting practitioner training and experience is clearly emphasized in the revised STRICTA Recommendations \([48]\). Five items from this group (questions nos. 12–16) pertain to aspects of the randomization and blinding procedures. As recommended in CONSORT, transparent reporting of these items is necessary to ascertain that selection and performance biases have been reduced \([17]\). A further low-scoring question (no. 23) is that calling for reporting of adverse events. Data on this topic is of general importance since practitioner training and experience may influence the occurrence of adverse events \([34]\).

The mean OCSI score (63.0%; median 64%) of the 333 acupuncture RCTs included in the present review indicates a moderate level of adherence to the CONSORT and STRICTA guidelines. This score is somewhat higher than those from a study that assessed observance of STRICTA as well as a selected set of CONSORT recommendations in samplings of acupuncture RCTs appearing both before and after publication of each set of guidelines \([34]\). It is also of interest that our subset of 209 acupuncture RCTs that appeared in biomedical (in contrast to CAM) journals received a mean score of 69.9%, similar to the adherence levels to CONSORT in assessments of RCTs of psychotropic pharmaceuticals \([28]\), general medicine interventions \([50, 51]\), endocrinology \([23]\), and surgical procedures \([27, 52]\).

Although a significant trend toward an annual increase in reporting quality is apparent in our findings, the 68.7% mean OCSI score for the first three-quarters of 2007 (the most recent period evaluated) indicates a continuing need to improve reporting in acupuncture RCTs. As a secondary analysis, composite scores were examined of both the CONSORT- and STRICTA-based questions. Analysis of CONSORT-based questions reveal an improvement of 18% over the 10-year period \((P < .001, 95\% CI = 0.011–0.025)\), with STRICTA-based questions improving by 17% \((P < .001; 95\% CI = 0.006–0.017)\). While a similar increase in CONSORT-related reporting in acupuncture RCTs was observed by Prady et al. \([34]\), the STRICTA-related increase was not previously observed, a difference likely related to the inclusive (present study) rather than sampling (Prady et al.) approaches utilized. Although the increased trends are apparent, the overall mean OCSI score (63%) implies that reporting of the acupuncture intervention (STRICTA) needs improvement as much as reporting of general aspects of research design (CONSORT).

Interpretation of our results is impacted by the included RCTs from China, which comprise, when ranked by country, both the largest number of trials and the lowest mean OCSI score (Table 2). When RCTs from China are excluded, the mean OCSI score increases from 63.0 (16.5) to 68.4 (13.5). A matter of concern is that many RCTs from China either did not actually randomize participants or utilized inadequate procedures \([53]\). It is likely that this finding adversely impacted, the OCSI score of RCTs from China. Although, over the decade examined the OCSI scores from China are consistently lower than those from all other countries, the translations of both CONSORT and STRICTA into Chinese should help to close the gap in quality of reporting \([44]\).

A recent national effort in China to improve trial design of acupuncture RCTs should also contribute to enhanced reporting quality \([54]\).

Two further cautionary notes apply when considering interpretations of OCSI scores. First is the somewhat surprising finding that 54% of authors responding to a survey on their use of STRICTA reported that some checklist items were removed in the journal review process in response to editorial concern regarding space constraints \([47]\). If this is indeed a widespread occurrence, journals should be encouraged to request alternate means for authors to provide STRICTA and CONSORT details, for example, electronic links to an expanded methods section. Second, in calling for acupuncture RCTs to meet reporting standards set by CONSORT and STRICTA, it should be considered that many acupuncture trials have been and continue to be designed as “early phase” research \([55]\). This does not mean that such trials should be allowed a lower standard of reporting quality but does indicate that the standards should be applied in a manner appropriate to the early stage of the research, when issues concerning patient recruitment, compliance, and retention are as critical to assess as are “hint of efficacy” and maintenance of benefit. For example, an early phase study should not “lose credit” for lack of reporting a sample size calculation if, in fact, no preliminary data existed on which to base such a calculation. Accordingly, in this case, the OCSI group decided that full credit should be awarded if a rationale was provided for the omission of a sample size calculation.

In addition to the limitations of our findings discussed above, there are several areas to consider for improving the OCSI instrument itself. For example, our decision to keep the numbering, wording, and grouping of the questions similar to those of the parent guidelines should be reconsidered since it may be of advantage to split out some of the sub-questions into stand-alone questions. One example would be to track, as a separate full question, whether the success of blinding study participants was evaluated. In the current OCSI format, evaluation of blinding is scored as a sub-question
grouped with related sub-questions that ask whether study participants, practitioners, and assessors were reported as blinded (question no. 16). While this grouping is appropriate as a single CONSORT guideline, a greater impact on overall scoring might be achieved if this sub-item was a separate question.

One can also argue that the comprehensiveness of OCSI is unnecessary to identify flaws in reporting and we should seek instead to identify a core set of questions that produce a quality score sufficiently similar to the total OCSI score. A potentially fruitful approach would be to survey stakeholders in the quality of reporting arena for sake of creating a weighted system for questions. Such an approach would minimize the “fatal-flaw” scenario of the unweighted instrument wherein a high score is achieved despite failure to report an item generally regarded as essential to “good reporting.” For this initial creation and application of OCSI, however, our goal was not to identify those items most essential for rigorous research design but to provide a means to assess comprehensive reporting in acupuncture RCTs, consistent with the aims of CONSORT and STRICTA.

Finally, in regard to application of the present findings, it is important to stress that quality of reporting is only a first step toward assessing the evidence base. For example, receiving high scores for reporting interventions may be of questionable clinical relevance if what is being reported are inad-

(ii) When multiple components (e.g., (i), (ii), (iii)) or “and” statements are listed for a question, all aspects must be met to receive full credit.

(iii) Except in regard to question no. 1, if the information is placed ANYWHERE in the document, the question can receive full credit.

(iv) For questions 4–9, which refer to STRICTA, nos. 4–8 pertain ONLY to the acupuncture aspects of the study, while no. 9 pertains to all types of control or comparison groups.

(v) Practitioners are almost always NOT BLINDED in acupuncture research; hence no. 16(b) is usually scored “N/A” (see no. 16(b) for further details).

A. Oregon CONSORT-STRIC TA
Instrument (OCSI) Manual

A.1. General Instructions

(i) When scoring each question consider the following.

(a) If you were a reviewer of the paper, would you be satisfied with what is reported?
(b) If you were attempting to reproduce the findings, is there sufficient reporting of details to allow you to do so?

No. 1. It must state in the abstract or title that participants were allocated to interventions in a randomized manner, that is, “random allocation,” “randomized,” or “randomly assigned.”

No. 2(a). “Scientific background” pertains to (i) the previous research and (ii) the current understanding of the disease.

No. 2(b). “Rationale” pertains to the justification for the trial and/or the expected effects/benefits of the treatment.

No. 3(a). The distinction between inclusion and exclusion criteria is unnecessary; for example, an exclusion criterion may be listed as a negative within the inclusion criteria. Stating the patient demographics and clinical characteristics after recruitment (e.g., post hoc demographics) does not qualify as eligibility criteria.

No. 3(b). “Settings,” for example, private practice or outpatient clinic and “Locations,” for example, medical center or acupuncture college, can be implied from the author affiliation information.

No. 4(a). “Style of acupuncture” can be implied from the author affiliation information; for example, authors were from the Beijing TCM hospital, so style is TCM. If the protocol utilized a single acupoint or only lists the names of the acupoint(s), the style still needs to be stated.

No. 4(b). For example, the acupoints were chosen on the basis of text books, practitioner survey, or prior research.

No. 4(c). “Justified,” for a YES score, requires citing journals, texts, practitioner survey, or an expert panel.

No. 5(a). Unilateral or bilateral can be implied.

No. 5(b). If the treatment protocol is individualized, then the authors must state the range of needles utilized (e.g., 8–14 needles). This may be implied from no. 5(a).

No. 5(c). If the treatments were individualized, the depths must be given as a range. Depths can be expressed in Cun, mm, or anatomical depth (e.g., subcutaneous tissue or muscle).

No. 5(d). “Response” refers to needle sensations such as, de qi (TCM), muscle twitch (trigger point), or muscle contraction (electroacupuncture).
No. 5(e). “Stimulation” refers to techniques in manual and electroacupuncture, for manual acupuncture: lifting, thrusting, or rotating; for electroacupuncture: the current, frequency, and amplitude.

No. 5(f). If the treatments were individualized “retention time” must be reported as a mean and range.

No. 5(g). It must state the gauge, length, and material and/or the manufacturer of the needles.

No. 6(a). If treatments were individualized, then the mean and range of the number of treatments must be stated.

No. 6(b). If the protocol was one treatment, then the frequency is “N/A”.

No. 7. It refers only to cointerventions within the scope of acupuncture licensure. If the protocol specifies the option of prescribed self-help treatments, for example, exercise advice, stretching, or qi gong, these must also be described.

No. 8(a). “Duration of practitioner training” can be implied, for example, LAc or a similar indication of formal training, (e.g., German Physicians Society of Acupuncture).

No. 8(b). “Clinical experience” refers to number of years in practice as an acupuncturist.

No. 8(c). “Expertise” must include experience, as an acupuncturist, with the condition under investigation.

Nos. 9(a–d). A no-treatment control is to be scored in the same manner as a sham or comparison control.

No. 9(a). “Intended effect” refers to the specific rationale for choosing the control (including no treatment), in relation to the research question and methodology.

No. 9(b). “Specific explanations” refers to the specific wording the patient receives regarding the treatment and control procedures. This information is needed, for example, since describing a sham procedure as “a type of acupuncture” may affect outcomes differently than stating “it isn’t acupuncture.”

No. 9(c). “Details” refers to a precise description of the control intervention including (if applicable, e.g., invasive sham) the parameters of needling (nos. 5(a–g)). This includes “no treatment” controls, for example, who may have been required to keep diaries. If no tracking was performed, it must be stated.

No. 9(d). “Justification” refers to citing journals, textbooks, practitioner survey, or an expert consensus panel.

No. 10(a). If objectives are clearly implied but not “specifically stated,” a “Yes” may be scored.

No. 10(b). The actual word “hypothesis” or a valid substitute, for example, theory, must be used.

No. 11(a). It must explicitly state (i) the primary outcome(s) and (ii) how it (they) was measured. If the primary outcome was not specifically indicated, partial credit can be awarded. (The primary outcome is usually used in the sample size calculation.)

No. 11(b). If particular steps were taken to increase the reliability of the measurements, sufficient details of the steps must be presented; for example, range of motion was assessed with a goniometer, and three successive measurements were averaged to increase reliability.

No. 12(a). It must include statements regarding estimated effect size of the intervention based on (i) previous research and (ii) estimated attrition rate. If the study was a pilot then a statement such as “No data exists upon which to base a sample size calculation” must be included.

No. 12(b). “Interim analysis and stopping rules” pertains to long-term trials and may not be applicable to trials of shorter duration. If authors report that interim analyses were performed, the number of assessments and statistical tests must be reported.

No. 13(a). Sufficient and specific information must be provided regarding the methods utilized to generate the random allocation scheme, for example, a computer-generated table. Simply stating that patients were randomly assigned is not sufficient to determine the likelihood of bias.

No. 13(b). If randomization restrictions were utilized, for example, stratification or blocking, the methods and details must be stated.

No. 14(a). “Implementation” refers to how the random assignment was applied; for example, a sealed envelope containing the computer generated-group assignment was placed in the patient’s chart.

No. 14(b). Concealment is required to prevent foreknowledge of and introduction of bias in treatment assignment, for example, the sealed envelope with group assignment was opened after the patient passed the baseline neurological exam and received a TCM differential diagnosis.

No. 15(a). It must specifically state the person responsible for generating the randomization.

No. 15(b). It refers to the specific person(s) responsible for consenting the participants.

No. 15(c). It refers to the person who assigned participants to their respective randomized groups. This may likely refer to the person who informed practitioners, directly or indirectly, of the participant’s group assignments. “Indirectly” may involve the preplacement of the group assignment in the patient chart. If “this” person also generated the randomization scheme, the authors should report how the potential for biased group assignments was addressed (e.g., randomization schedule was kept in a locked closet from “this” person).

No. 16(a). It must state that patients were blinded to group assignments, for example, a statement that the trial was single-blinded. If the trial was a comparison of active treatments, for example, acupuncture versus massage, this will be scored N/A.

No. 16(b). This will usually be scored N/A. In the rare case that the study did blind the practitioners, there must be an explicit statement regarding the method utilized.
No. 16(c). It must state whether or not the persons performing outcome measures were blinded to group assignments. This is more important with subjective outcomes, less so with objective outcomes where there is less chance or bias.

No. 16(d). For full credit, authors must (i) state that the success of participant blinding was evaluated and (ii) present the results of the assessment. This is N/A if acupuncture was compared to another active treatment (e.g., standard care or massage).

No. 17(a). It must specifically state the statistical methods used to compare groups for primary outcomes.

No. 17(b). If additional analyses were performed, the specific statistical methods used must be reported for full credit.

No. 18(a). A diagram is recommended but not necessary. For full credit, authors must state the number of participants who were (i) evaluated for potential enrollment, for example, excluded due to inclusion/exclusions criteria and declined enrollment (ii) randomized, (iii) received treatment, (iv) completed treatment, and (v) analyzed for the primary outcome.

No. 18(b). “Protocol deviations” pertains to the nature of protocol deviations after randomization, for example, departures from the protocol including unplanned changes to the intervention, data collection, examinations, and methods of analysis, or the specific reasons for patient exclusions after randomization, including drop-outs.

No. 19(a). It must state the start and finish dates of the recruitment period, to place the study in historical context.

No. 19(b). It must state the length of time for the followup period; for example, participants were followed for a total of 6 months, 3 months during treatment and 3 months of posttreatment followup.

No. 20(a). It must present the demographics, for example, age, gender, ethnicity.

No. 20(b). It must present the clinical characteristics, for example, duration of condition, for each group.

No. 21(a). For full credit, it must state the “n” per each analysis, since the actual “n” may vary per outcome. If no dropouts occurred, he “n” can be implied.

No. 21(b). Must use an “intention to treat” analysis (or a valid substitute, e.g., per protocol or on treatment analysis) to avoid bias associated with non-random loss of participants. If there were no dropouts then a “Yes” is scored.

No. 21(c). It must state absolute numbers, for example, reduction of VAS pain from 10 to 5 or from 6 to 3 not just a 50% reduction.

No. 22(a). It must state the results or use a table/graph, for example, the mean and standard deviation of the measurements.

No. 22(b). All results must indicate the effect size, for example, the difference in means, odds ratio, or relative risk ratio.

No. 22(c). For full credit, authors must state precision as a confidence interval (CI). Only presenting P-values will score a “partial.”

No. 23. If additional analyses are prespecified in the research design, the results must be presented, and it must be clearly stated whether or not these were exploratory (e.g., post hoc) analyses.

No. 24. It cannot be scored as N/A. Authors need to state whether there were or were not adverse events or side effects for each group. If severe AEs occurred, estimates of the frequency should be reported.

No. 25(a). The relevance of the findings to the hypotheses or objectives must be discussed.

No. 25(b). The strengths and weakness of the trial must be discussed, particularly in the context of potential for bias and imprecision of study findings.

No. 25(c). This is usually scored as N/A unless numerous analyses both prespecified and exploratory were performed.

No. 26. “Generalizability” pertains to the external validity or general applicability of the results. For example, to what extent do the results of the study, with the population limited by the eligibility criteria, pertain to the greater population with this condition? Topics to consider are eligibility criteria, setting and location of the study, the intervention protocol, the period of recruitment and followup, and so forth.

No. 27. “General interpretation” pertains to how these results relate to previous research. For full credit, authors must (i) cite and (ii) discuss the previous research.

B. Reference List of 333 RCTs Included in OCSI (Alphabetical)


(111) Jiang H, Shi K, Li X, Zhou W, Cao Y. Clinical study on the wrist-ankle acupuncture treatment for 30 cases of...


(190) Nedstrand E, Wyon Y, Hammar M, Wijma K. Psychological well-being improves in women with breast cancer after treatment with applied relaxation


and vomiting following pediatric tonsillectomy with or without adenoidectomy. *Anesthesiology* 2002; 96:300–305.


for Pain Relief after Ambulatory Knee Arthroscopy—


(271) Vincent A, Barton DL, Mandrekar JN, Cha SS, Zais Waite NR, Clough JB. A single-blind, placebo-

(272) Wang B, La J, Wang BX. Clinical Studies on sci-


(275) Wang L. Clinical observation on acupuncture treat-

(276) Wang L. Clinical observation on acupuncture treat-

(277) Wang RR, Tronnier V. Effect of acupuncture on pain manage-


(284) Wedenberg K, Moen B, Norling A. A prospective random-
ized study comparing acupuncture with physio-


(321) Zaborowska E, Brynhildsen J, Damberg S, Fredriksen M, Lindh-Astrand L, Nedstrand E et al. Effects of acupuncture, applied relaxation, estrogens and


Acknowledgments

Many thanks to Liz Collins and Laura Varga for their contributions as OCSI raters. Helpful feedback on the manuscript from Hugh MacPherson and Peter Wayne is appreciated. This study received personnel support from the research department of the Oregon College of Oriental Medicine and the Helfgott Research Institute of the National College of Natural Medicine.

References


Submit your manuscripts at http://www.hindawi.com