Evaluating the Quality of Reporting Occupational Therapy Randomized Controlled Trials by Expanding the CONSORT Criteria

Emily Moberg-Mogren, David L. Nelson

CONTEXT. Well-reported randomized controlled trials (RCTs) are necessary for evidenced-based practice in the field of occupational therapy.

OBJECTIVE. The primary purpose of this study was to analyze the quality of reporting in selected RCTs relevant to occupational therapy by using the Nelson-Moberg Expanded CONSORT Instrument (NMECI). The interrater reliability of the NMECI was also tested.

METHOD. The 201-sub-item NMECI was developed by the authors to clarify the 22-item CONSORT Statement, each item of which raises several issues. Fourteen RCTs obtained through the OTseeker database were analyzed independently by the authors.

RESULTS. The 14 articles complied with slightly more than half of the sub-items of the NMECI ($M = 104.2$, $SD = 32.9$ for the main rater). The only area of full compliance involved reporting of objectives and hypotheses; areas of frequent noncompliance included reporting of adverse events, randomization, and blinding. The intraclass correlation coefficient assessing level of interrater agreement on the total NMECI score was .95, indicating strong agreement overall. For the 176 kappas that could be computed on individual sub-items, the median kappa was .85, with an inter-quartile range from .58 to 1, indicating high levels of agreement for most items. However, 29 kappas fell below the moderate level of agreement (.40).

CONCLUSIONS. The field of occupational therapy needs increased sophistication in conducting and reporting RCTs. The NMECI shows promise in evaluating the quality of RCTs, given refinements to some items.

Kielhofner, Hammel, Finlayson, Helfrich, and Taylor (2004) have posited that research on outcomes is essential to the profession of occupational therapy. They went on to argue that several different research methodologies are needed to study occupational therapy outcomes. Specifically they described the potential contributions of measurement studies, database research, qualitative research, meta-analyses, and randomized controlled trials (also called randomized clinical trials, or RCTs). Of these designs, the RCT “most rigorously ensures that outcomes can be attributed to treatment versus other intervention effects” (p. 18). Although it is important to note that RCTs are not appropriate for all occupational therapy research questions, it is also important to recognize the broad and stable consensus among scientists that RCTs provide the most definitive evidence of efficacy and effectiveness in health care (Altman et al., 2002; National Institutes of Health, Division of Research Grants, Research Analysis and Evaluation Branch, 1979).

Essentially, an RCT is a refined experiment, with random assignment of participants to different interventions (and/or to control conditions), with systematic testing of outcomes, and with controls over possible confounding variables (Friedman, Furberg, & DeMets, 1998, Chap. 1). The typical problems of studying health care have led to additional refinements in RCTs over the years, including special procedures for randomizing participants (often new patients) entering the study at various times and for blinding (masking) researchers and participants.
to prevent possible bias (Moher, Schulz, & Altman, 2001).

In the 2000 Eleanor Clarke Slagle lecture, Holm (2000) argued that the best support for evidenced-based occupational therapy practice consists of systematic reviews of multiple, well-designed RCTs. According to Holm, the profession must foster increased knowledge among therapists of RCTs, including guidelines used in judging quality. According to the American Occupational Therapy Association (AOTA) Standards of Practice (American Occupational Therapy Association [AOTA], 1998), the occupational therapist must be able to apply research findings ethically and appropriately to evaluation and intervention procedures. This competency clearly depends on the ability to evaluate the quality of research articles, particularly RCTs.

In a recent paper, Stern (2005, p. 161) posed the following questions to her research students: “How well written were the articles? How do you decide for yourself whether an article is well written?” To answer these questions, students, practitioners, and instructors need detailed guidelines refined over time within the scientific community. Quality of reporting directly affects the reader’s ability to evaluate the validity and potential usefulness of a research article: “Critical appraisal of the quality of clinical trials is possible only if the design, conduct, and analysis of RCTs are thoroughly and accurately described in published articles” (Altman et al., 2002, p. 1).

To address the issue of adequate reporting of research, international journal editors, trialists, and methodologists developed the CONSORT (Consolidated Standards of Reporting Trials) Statement (Begg et al., 1996). The current, revised version of the CONSORT provides a checklist of 22 essential items that should be included in an RCT, and a diagram for documenting the flow of participants through a trial (Moher, Shulz, & Altman, 2001). Prominent features of the CONSORT include an emphasis on randomization and detailed reports of participant selection, allocation, and methods. Numerous medical and health care journals and organizations have adopted the CONSORT, available on the Internet (http://consort-statement.org). For example, editors of journals oriented to complementary medicine have accepted CONSORT and then built upon it by adding specific guidelines for acupuncture called STRICTA (Standards for Reporting Interventions in Controlled Trials of Acupuncture) (MacPherson et al., 2001).

Of course, research reporting in occupational therapy is somewhat different from reporting drug trials in medicine (Nelson & Mathiowetz, 2004). For instance, often there is little previous research on which to build in many areas of occupational therapy practice (CONSORT item 1). It is also difficult to describe all aspects of interventions, such as the therapeutic relationship (CONSORT item 4). Furthermore, it is usually impossible to mask participants and interventionists in occupational therapy research (CONSORT item 11). Nevertheless, it is essential that occupational therapy RCTs move toward compliance with CONSORT criteria in all possible areas. There is no reason why occupational therapy RCTs cannot describe participants, outcome measurements, randomization procedures, and statistical procedures as detailed in CONSORT.

A problem in using the CONSORT is that each of the 22 items consists of multiple statements, each of which can be evaluated independently. For example, CONSORT item 1 requires that (a) the term random be used in the title, (b) the term random be used in the abstract, and (c) the abstract be structured by subheadings. This format is adequate for authors and editors, but problematic for quantitative use of CONSORT. One solution would be to require all components within the item in order to certify the item as “passed.” However, this would equate an article that complied with none of the components with an article that complied with all but one. Another solution has been to score each CONSORT item as met, partially met, or not met (Herbison, 2005). A third solution has been to divide each CONSORT item into sub-items. Hence Stinson, McGrath, and Yamada (2003) rated articles on 43 items drawn from CONSORT, and Moher, Jones, and Lepage for the CONSORT Group (2001) used 40 items. We reviewed and tested these latter two models, but found that the problem of multiple issues per item remained. Therefore, we developed a rating scale based exclusively on the CONSORT but consisting of 201 sub-items.

The primary purpose of this study was to apply this detailed interpretation of the CONSORT criteria, the Nelson-Moberg Expanded CONSORT Instrument (NMECI), to RCTs relevant to occupational therapy. To what extent do reported occupational therapy RCTs comply with detailed CONSORT Statement criteria? This study also tested the interrater reliability of the NMECI.

Method

Preliminary Expansion of the CONSORT Statement

The reason for expanding the 22-item CONSORT Statement was that each item consisted of multiple, independent concepts, each of which can be evaluated. Based on the elaboration document, The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration (Altman et al., 2002), the first author, EM, systematically divided each of the 22 CONSORT items into a...
more detailed checklist of 143 sub-items. The raters (DN and EM) conducted practice analyses on two published RCTs not used in the final analysis. The articles were from the fields of psychology and occupational therapy (Esplen, Garfinkel, Olmsted, Gallop, & Kennedy, 1998; Mann, Ottenbacher, Fraas, Tomita, & Granger, 1999). The raters met to analyze the two practice articles according to the 143-sub-item checklist so that the criteria could be discussed, applied, and analyzed in terms of potential problems of agreement in applying the criteria. Throughout the training, the raters kept a list identifying possible ambiguities in applying the criteria. DN used this list and additional review of the elaboration document to create the 201-sub-item NMECI. Both raters then used the final instrument to practice the rating system on three medically oriented articles unrelated to occupational therapy.

**Rules for Sub-Items of the Nelson-Moberg Expanded CONSORT Instrument**

Throughout the process of expanding the CONSORT, the authors frequently encountered problems of logic, validity, and interpretability. To overcome these problems, the following rules were developed for consistent expansion of the CONSORT Statement into sub-items:

1. Generally, only one issue of research design or reporting can appear in each Nelson-Moberg sub-item. For example, reporting on the location of recruitment is a separate issue (and a separate sub-item) from reporting on the method of recruitment. The only exception to this rule occurs when an issue is variable (one or more). For example, a study might involve one secondary outcome or might involve several. In order to have the same number of sub-items to evaluate all articles, the decision was made to evaluate each variable issue in an all-or-none way (e.g., if the properties of any of the secondary outcomes were not described, no credit was granted for that sub-item).

2. Except for four sub-items (see 3 below), each item is rated as a 1 (compliance with CONSORT) or 0.

3. Four sub-items in CONSORT item 21 are scored as percentages. CONSORT item 21, dealing with generalizability, refers to CONSORT items 3, 4, 6, and 14, each of which has several Nelson-Moberg sub-items. Therefore, the score for generalizability includes sub-items reflecting the percentages received for CONSORT items 3, 4, 6, and 14. For example, if a study received 11 of 13 possible for CONSORT item 3, then a score of .85 would be given to the corresponding Nelson-Moberg sub-item.

4. Each NMECI sub-item involves an issue that can be scored as present or absent by a reviewer who is not an expert in the particular field under study and who is not a member of the research team. An example of an issue that cannot be scored by a nonexpert is in CONSORT item 2, where authors are urged to report the scientific background for the study. A nonexpert reviewer cannot evaluate the completeness or fairness of the authors’ descriptions of scientific background; however, a reviewer can judge whether scientific background was given or not. An example of an issue that cannot be known by someone who was not a part of the research team is the CONSORT exhortation to report deviations from the research protocol. However, the reviewer can make a judgment as to reporting of whether or not deviations from the research protocol occurred (with follow-up explanations if they did occur).

5. If the same CONSORT issue appears in more than one CONSORT item, it is scored separately in each section. For example, CONSORT items 20 and 22 both call for a systematic review. Therefore, a study reporting a systematic review can receive credit on a Nelson-Moberg sub-item within CONSORT item 20 and can also receive credit on a Nelson-Moberg sub-item within CONSORT item 22.

6. Nelson-Moberg sub-items reflect the ideals cited by CONSORT, not the minimal requirements. For example, CONSORT item 13 states that a diagram of participant flow is strongly recommended; therefore, an NMECI sub-item deals with this issue.

7. RCTs that do not comply with CONSORT rules calling for clear identification of primary versus secondary outcomes (and which therefore are not credited on the relevant Nelson-Moberg sub-item) are not further penalized in subsequent sub-items dealing with the outcomes. Hence, an RCT that reports the provenance of all its outcomes receives credit for Nelson-Moberg sub-items dealing with reporting the provenance of primary and secondary outcomes.

8. Credit is granted if an NMECI sub-item involves an issue that is irrelevant to the RCT, but only if certain basic information in that CONSORT item is provided. For example, in CONSORT item 8, Nelson-Moberg sub-items dealing with full reporting of blocking and stratification can be earned if blocking and stratification were not done in the study, but only if a basic description of the randomization process is given.

9. Generally credit is granted even if the authors provide the information in a different section of the paper. For example, credit for describing the participants (CONSORT item 3, dealing with the Method section) is granted if the authors describe the participants in the results. The
exception to this rule is that the scientific background and explanation of rationale (CONSORT item 2) must be given prior to the Method section (e.g., not in the Discussion section). This rule is consistent with the CONSORT elaboration.

An example of 13 NMECI sub-items drawn from CONSORT item 3 is provided in Figure 1. Also given are rules of interpretation consistent with the nine rules stated above.

**Selection of the Articles to be Analyzed**

RCTs were selected through OTseeker: Occupational Therapy Systematic Evaluation of Evidence, a database that includes meta-analyses and randomized trials pertaining to occupational therapy (OTseeker, 2003). In order for an RCT to be included in the study, (1) the trials had to be locatable using a search of OTseeker limited to controlled trials with “occupational therapy” in the title of the article; (2) the RCT included random allocation of at least two interventions (or one control condition compared to one or more interventions); (3) each study participant received only one condition of the independent variable (eliminating counterbalanced/crossover designs); (4) at least one of the interventions was part of occupational therapy practice, or could become part of occupational therapy practice; (5) the dependent variable involved an outcome of therapy, not a short-term indicator of motivation or some aspect of the therapeutic process such as exercise repetitions; (6) the trial involved human participants; (7) the trial was reported as a full-size article in a peer-reviewed journal; and (8) only one study per sample could be included (thus eliminating follow-up and related studies of previously reported participants). No limits were placed on dates of publication.

Twenty-one studies were found in OTseeker on November 5, 2003, using the limitation of controlled trials and the key phrase “occupational therapy.” One study was then excluded because it used a counterbalanced design, three studies were excluded because they did not involve therapy outcomes, and three studies were excluded because they used the same sample as another study to report follow-up or related results. The 14 articles are indicated by * in the reference list.

---

**Table: Nelson-Moberg Expanded CONSORT Instrument Elaboration Sub-Items Based on CONSORT Explanation Document**

<table>
<thead>
<tr>
<th>Section</th>
<th>Topic</th>
<th>Item</th>
<th>Nelson-Moberg Expanded CONSORT Instrument Elaboration Sub-Items Based on CONSORT Explanation Document</th>
<th>Interpretation/Explanation</th>
</tr>
</thead>
</table>
| METHODS | Participants | 3 A. Eligibility criteria for participants | 11) ______________ define eligibility criteria  
12) ______________ describe the method of recruitment (e.g., referral or self-selection through advertisements)  
13) ______________ describe the geographical location(s) for recruitment (e.g., city/country as needed)  
14) ______________ describe the type(s) of setting, (e.g., community, office practice, hospital clinic, hospital)  
15) ______________ describe the trial participants | A. overlaps with part of CONSORT 15, and is scored in both sections.  
(11) need not distinguish between exclusion and inclusion criteria  
(13) if used, advertisements must imply the geographical area (e.g., by identifying the publications or posting sites)  
(15) may be reported in Results section (overlap with CONSORT # 15) |
|         |       | 16) ______________ report the number of intervention settings  
17) ______________ report the number of data collection sites  
18) ______________ report the types of settings where intervention took place (e.g., community, office practice, hospital clinic, hospital inpatient)  
19) ______________ report the geographical locations of the intervention sites (e.g., city/country)  
20) ______________ report the type(s) of settings where data collection took place  
21) ______________ report the geographical locations of the data collection sites  
22) ______________ report the number of care providers  
23) ______________ report the types of care providers (e.g., credentials) | A. (13)–(14), B. (18)–(21) may be reported in one place if the settings were same  
(16)–(17) credit if clearly only one site  
(22)–(23) credit if the only care providers are the interventionists and if the interventionists are described in CONSORT # 4 (27)–(28) |
|         |       | Sub-total __________ / 5 |
|         |       | Sub-total __________ / 8 |
|         |       | Total _____________ /13 |

**Figure 1. Nelson-Moberg Expanded CONSORT Instrument elaboration of CONSORT item #3 into 13 sub-items.**
After identifying the 14 RCTs, the evaluators independently used NMECI to systematically analyze the articles from June 2004 to September 2004. The evaluators had no contact with each other during the evaluation.

Plan for Data Analysis

The main indicator of compliance with CONSORT was planned as the average (mean or median, depending on skewness) NMECI score when all of the 201 sub-items were summed (a possible range from 0 to 201). Also planned was computation of average compliance scores for each of the 22 CONSORT items and each of the 201 sub-items. DN’s scores were used as primary data, with EM’s scores used for interrater reliability purposes. The reason for using DN’s scores was his relative level of research experience.

The main test for interrater reliability dealt with the source of the main indicator of compliance, the total score. The intraclass correlation coefficient for generalization to a future single rater was used, formula 2.1, as described in Shrout and Fleiss (1979). The same statistic was used to evaluate each of the 22 CONSORT items. Simple kappa was then used test interrater agreement on 197 of the 201 items, with four of the items (sub-items 193–196) tested by the ICC because they involved percentages, not categorical judgments. We used the guidelines proposed by Landis and Koch (1977) for interpreting the strength of kappa and guidelines proposed by Eliasziw, Young, Woodbury, and Fryday-Field (1994) for interpreting the ICC.

Results

Interrater Reliability

The main test of agreement yielded an ICC of .95, indicating “almost perfect” strength of agreement (Eliasziw et al., 1994). Next an ICC was computed (testing the level of agreement on the Nelson-Moberg sub-items reflecting each CONSORT item). ICC results for each of the 22 CONSORT items are presented in Table 1. This table indicates perfect or “almost perfect” strength of agreement for items 1, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 20, and 21; “substantial” strength of agreement for items 2 and 3; “moderate” strength of agreement for item 22; “fair” agreement for item 11; “slight” strength of agreement for item 18; and “poor” strength of agreement for item 19.

ICCs were computed for sub-items 193 (ICC = .43), 194 (ICC = .83), 195 (ICC = .59), and 196 (ICC = 1). In addition, kappa values could not be computed for 21 sub-items because these sub-items had no variance (there was perfect agreement that either all 14 articles met the criterion, Table 1. Mean Compliance With CONSORT Items and Overall CONSORT (Based on DN’s Ratings), and Intraclass Correlation Coefficients Reflecting Levels of Interrater Agreement

<table>
<thead>
<tr>
<th>Section/Topic</th>
<th>CONSORT Item</th>
<th>NMECI Sub-Items</th>
<th>Mean</th>
<th>SD</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title/Abstract</td>
<td>1</td>
<td>1–3</td>
<td>1.79</td>
<td>1.12</td>
<td>1.00</td>
</tr>
<tr>
<td>Background</td>
<td>2</td>
<td>4–10</td>
<td>3.71</td>
<td>1.59</td>
<td>.78</td>
</tr>
<tr>
<td>Participants</td>
<td>3</td>
<td>11–23</td>
<td>9.93</td>
<td>2.02</td>
<td>.73</td>
</tr>
<tr>
<td>Interventions</td>
<td>4</td>
<td>24–30</td>
<td>3.50</td>
<td>1.56</td>
<td>.88</td>
</tr>
<tr>
<td>Objectives</td>
<td>5</td>
<td>31–32</td>
<td>2.00</td>
<td>0.00</td>
<td>—d</td>
</tr>
<tr>
<td>Outcomes</td>
<td>6</td>
<td>33–52</td>
<td>13.00</td>
<td>3.98</td>
<td>.90</td>
</tr>
<tr>
<td>Sample size</td>
<td>7</td>
<td>53–66</td>
<td>6.64</td>
<td>3.41</td>
<td>.95</td>
</tr>
<tr>
<td>Randomizationa</td>
<td>8</td>
<td>67–79</td>
<td>8.00</td>
<td>5.46</td>
<td>.96</td>
</tr>
<tr>
<td>Randomizationb</td>
<td>9</td>
<td>80–83</td>
<td>0.43</td>
<td>1.16</td>
<td>1.00</td>
</tr>
<tr>
<td>Randomizationc</td>
<td>10</td>
<td>84–89</td>
<td>0.57</td>
<td>1.09</td>
<td>1.00</td>
</tr>
<tr>
<td>Blinding</td>
<td>11</td>
<td>90–102</td>
<td>1.07</td>
<td>0.92</td>
<td>.31</td>
</tr>
<tr>
<td>Statistics</td>
<td>12</td>
<td>103–115</td>
<td>4.93</td>
<td>4.80</td>
<td>.88</td>
</tr>
<tr>
<td>Participant flow</td>
<td>13</td>
<td>116–132</td>
<td>10.50</td>
<td>5.75</td>
<td>.98</td>
</tr>
<tr>
<td>Recruitment</td>
<td>14</td>
<td>133–138</td>
<td>3.14</td>
<td>1.88</td>
<td>.98</td>
</tr>
<tr>
<td>Baseline data</td>
<td>15</td>
<td>139–145</td>
<td>4.14</td>
<td>2.57</td>
<td>.91</td>
</tr>
<tr>
<td>Numbers analyzed</td>
<td>16</td>
<td>146–153</td>
<td>6.29</td>
<td>2.49</td>
<td>.96</td>
</tr>
<tr>
<td>Outcomes</td>
<td>17</td>
<td>154–166</td>
<td>6.93</td>
<td>5.14</td>
<td>.95</td>
</tr>
<tr>
<td>Ancillary analyses</td>
<td>18</td>
<td>167–175</td>
<td>8.57</td>
<td>0.85</td>
<td>.12</td>
</tr>
<tr>
<td>Adverse events</td>
<td>19</td>
<td>176–180</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Interpretation</td>
<td>20</td>
<td>181–189</td>
<td>5.00</td>
<td>1.66</td>
<td>.88</td>
</tr>
<tr>
<td>Generalizability</td>
<td>21</td>
<td>190–197</td>
<td>3.08</td>
<td>1.33</td>
<td>.89</td>
</tr>
<tr>
<td>Overall evidence</td>
<td>22</td>
<td>198–201</td>
<td>1.00</td>
<td>1.24</td>
<td>.53</td>
</tr>
<tr>
<td>Overall Score</td>
<td>1–201</td>
<td>104.23</td>
<td>32.93</td>
<td>.95</td>
<td></td>
</tr>
</tbody>
</table>

---

Footnotes:

a Sequence generation
b Allocation concealment
c Implementation
d Perfect agreement that all articles met this criterion; ICC cannot be calculated (no variance).
or that none met the criterion). Please see Table 2. The mean $kappa$ on the remaining sub-items was .85, with an inter-quartile range from .58 to 1. According to Landis and Koch (1977), this indicates a high degree of agreement overall. Fully 126 of 176 $kappas$ were perfect or were in the substantial or almost perfect ranges, according to Landis and Koch. However, 29 of the 176 $kappas$ ranged from -.08 to .34, in the ranges termed poor, slight, and fair by Landis and Koch. Table 3 lists these sub-items and also includes information concerning percentage of agreement. Percentage of agreement is given because $kappa$ can be low or even 0 when the two raters agree at high rates. For example, when the two raters agree in 13 of 14 cases (93%) that the criterion is met, but disagree in the 14th case, $kappa$ is 0.

**Degree of Compliance of the Articles to CONSORT**

Overall, the articles met about half of the sub-items of the NMECI based on DN's ratings ($M = 104.2$, $SD = 32.9$, skewness = -.09). See Table 1 for the degree of compliance in each of the 22 CONSORT items. Eighty-one of the 201 NMECI sub-items were below a mean of .5, indicating that less than half of the articles met the criteria of these sub-items. Seven of the items were not met by any of the articles. For example, sub-item 29, which states, “describe the types of interventionists (duration of experience),” and sub-item 55, which states, “identify the estimated outcomes in each group (not just the expected difference between groups),” were not reported in any of the articles.

**Additional Analyses**

A $t$ test was computed to compare evaluators in terms of positive versus negative ratings. DN's ratings ($M = 104.2$, $SD = 32.9$) were significantly lower than EM's ratings ($M = 112.0$, $SD = 31.0$) ($t = 3.75$, $p = .002$). A moderate yet statistically significant positive correlation was found between year of publication and CONSORT compliance, ($rho = .55$, $p = .04$). Years of publication ranged from 1975 to 2001, with a median between 1996 and 1997.

**Discussion**

**Degree of Compliance of the Articles to CONSORT**

Independent raters agreed that RCTs relevant to occupational therapy published in peer-reviewed journals met slightly more than half (52% and 56%) of the requirements outlined by the CONSORT criteria. A few of the criteria identified in the CONSORT Statement are impossible to comply with in most occupational therapy research, such as blinding of interventionists and subjects (NMECI sub-items within CONSORT item 11). However, most of the sub-items are possible to report as recommended by CONSORT. Full reporting helps the reader of
the article to make decisions about the internal and external validity of the study, ultimately leading to improved clinical decision making.

CONSORT item 5 (stating the specific objectives and hypotheses) was the only item judged to be in full compliance by all 14 articles. Other areas of relative strength include CONSORT items 3 (reporting characteristics of participants) and 16 (reporting number of participants analyzed). Relatively weak areas were CONSORT items 9 (methods to conceal randomization), 10 (methods to implement randomization), 11 (blinding), and 19 (description of adverse events). These are areas that are not particularly difficult to comply with in occupational therapy research. For example, although it is impossible to blind interventionists and participants in most occupational therapy research, it is easy to report that they were not blinded.

We suggest that researchers, reviewers, and editors develop the habit of addressing these issues. For example, CONSORT will be complied with if it is simply stated that no adverse events occurred, if that is the case. It is interesting to note that a recent CONSORT Group initiative has resulted in new guidelines for reporting adverse events, as a supplement to the main CONSORT criteria (Moher, Altman, Schulz, & Elbourne, 2004).

The ratings in this report should not be interpreted as a rationale for disregarding the results of the research reported in these 14 articles. Many of the articles used in this study were written before publication of CONSORT and its adoption in peer-reviewed journals. Because the articles analyzed in this study use experimental design with randomization, they are most likely among the strongest research studies conducted in the field of occupational therapy. In fact, six of the articles met over half of the criteria. One article (Walker et al., 1999) met 150.8 of the 201 sub-items, and another (Parker et al., 2001) met 146.1 of the 201 sub-items. It is also important to point out that parallel studies of adherence to CONSORT criteria have indicated parallel problems of reporting in fields as diverse as pediatric psychology (Stinson et al., 2003) and obstetrical anesthesia (Halpern, Darani, Douglas, Wight, & Yee, 2004). A particularly weak area across all fields has been the description of randomization procedures.

The quality of reporting is not the same thing as quality of research (Huwiler-Müntener, Jüni, Junker, & Egger, 2002). For example, a highly flawed study with a small sample could score quite high (but not perfectly) on the CONSORT or the NMECI as long as the reporting of the flaws was thorough. Conversely, an excellent trial might be reported incompletely because of the authors’ training or because of editorial conventions. A practical problem in reporting RCTs is that many journals, particularly those oriented to medical specialties, have stringent page requirements, resulting in loss of important detail. Specifically, introductions tend to be brief, and therefore the state of science in the area of concern and the theoretical background of the study tend to be underreported. Systematic study of the relationship between quality of reporting and quality of research is difficult, given the inherent problem of judging the quality of research that is poorly reported.

The moderately positive correlation between year and score on the NMECI suggests an improvement of RCT reporting in occupational therapy (the more recent the article, the higher the degree of CONSORT compliance). With the growth of the field, there has been an increase in the quality of research. As the field of occupational therapy expands and as investigative efforts increase, research results must be reported in such a way that practitioners and fellow researchers can determine their veracity (Holm, 2000). In order to provide the most transparent information about “why a study was undertaken and how it was conducted and analyzed” (Moher et al., 2001, p. 657), the CONSORT can be used to guide authors in writing thorough research reports forming a basis for evidence-based practice.

Development of the Nelson-Moberg Expanded CONSORT Instrument

When considering the overall score as the principal indicator of quality, the intrater reliability of the NMECI is excellent with “almost perfect strength of agreement.” This may have been due to the fact that the authors met before the analyses to discuss the meaning for each item based on The Revised CONSORT Statement for Reporting Randomized Trials: Explanation and Elaboration (Altman et al., 2002). The descriptions presented for each item also may have contributed to the high overall agreement. It is important to note that the CONSORT was not modified by the authors, but rather expanded into a more detailed checklist based solely on the CONSORT elaboration document. Moher et al. (2001) stated that an “iterative process” makes the CONSORT Statement a continually evolving instrument (p. 658).

However, 3 of the 22 ICCs reflecting agreement on the original CONSORT items and 29 of the kappas on NMECI sub-items were unacceptably low (fair, slight, or poor agreement). Sub-item 10, “Describe the broad approach taken to study the problem,” requires subjective analysis to receive credit. The term “broad” does not define what is necessary to report in describing the process of the intervention. In the future, this term could be more clearly
defined to increase the reliability of the item, by naming specific aspects to be included in the report, such as a general description of the intervention and participants. Subitem 15 also uses general terms, lacking specific instruction; hence raters relied on subjective judgments.

Sub-items 45, 46, 77, 189, and 197 may have had low kappas and poor agreement because of a difference in training and knowledge between the two raters in regard to statistical methods. A limitation of this study is that EM did not fully understand certain statistical procedures or methodologies. NMECI sub-items were taken directly from the CONSORT elaboration document, which assumes users’ prior knowledge of principles of research design.

Future Research

Although the total score of the NMECI proved to be highly reliable, sub-items with low kappa values should be revised prior to future research. In some cases, the wording of the sub-items should be changed to increase clarity. In other cases, the explanation of the sub-item should be amplified to cover previously unanticipated problems of interpretation. Also, RCT experts could offer systematic feedback on the content validity of the NMECI sub-items and the procedures used in scoring. Once refined, the NMECI could be used for analyzing the quality of reporting in occupational therapy RCTs not cataloged in the OTSeeker database with “occupational therapy” in the title; for comparing quality of reporting across fields; and for comparing quality of reporting across time in the future. The NMECI could also be used to investigate whether incompleteness of reporting is associated with large effect sizes in occupational therapy research, as was done by Stinson et al. (2003) in psychology RCTs. A final recommendation is the development of supplementary sub-items dealing with issues commonly arising in occupational therapy and related fields involving complex interactions between interventionists and research participants. For example, issues dealing with intervention fidelity differ greatly between drug studies and occupational therapy intervention studies (Nelson & Mathiowetz, 2004).

Implications for Occupational Therapy

The main significance of this study is that the reporting of occupational therapy RCTs can and should be improved in the future, in support of evidence-based practice. Tickle-Degnen (1999, p. 538) pointed out that evidence-based practice depends in part upon understanding research procedures and upon evaluating the validity of the study. Incompletely described procedures hamper the practitioner’s search for evidence by casting doubt on the research conclusions. For example, the practitioner cannot know to whom the results can be generalized if the sample is described incompletely. For another example, the practitioner cannot trust the findings if bias might have been introduced by randomization measurement procedures that are described incompletely. Incompletely described procedures leave the practitioner in a state of doubt. Does the lack of detail reflect design flaws or simply errors of writing? There is no way of knowing the validity of the procedures without knowing what the procedures were. This study alerts researchers as well as consumers of research to past problems and the need for future improvement.

The usefulness of a detailed interpretation of the CONSORT Statement such as the NMECI depends on one’s perspective. Those designing RCTs would be well advised to use such a list as a double-check to their memories. Also, authors have special access to information concerning a trial, but often struggle with editors as to inclusion of details. A detailed interpretation such as the NMECI could help authors and editors negotiate what is critical information, and what is not. This could be helpful even if a 201-sub-item checklist is judged too detailed for a standard review. Although it is unlikely that clinicians would have time to apply such a detailed analysis, it might be helpful to them if such an analysis was provided by a Web site. So much work goes into conducting an RCT that the extra time involved in reporting it well is worthwhile.

References

(14 studies analyzed in this paper are marked by *)


