Reporting of Randomized Controlled Trials in Hodgkin Lymphoma in Biomedical Journals

Thilo Kober, Sven Trelle, Andreas Engert

Background: Randomized controlled trials (RCTs) are the best tool to evaluate the effectiveness of clinical interventions. The Consolidated Standards for Reporting Trials (CONSORT) statement was introduced in 1996 to improve reporting of RCTs. We aimed to determine the extent of ambiguity and reporting quality as assessed by adherence to the CONSORT statement in published reports of RCTs involving patients with Hodgkin lymphoma from 1966 through 2002. Methods: We analyzed 242 published full-text reports of RCTs in patients with Hodgkin lymphoma. Quality of reporting was assessed using a 14-item questionnaire based on the CONSORT checklist. Reporting was studied in two pre-CONSORT periods (1966–1988 and 1989–1995) and one post-CONSORT period (1996–2002). Results: Only six of the 14 items were addressed in 75% or more of the studies in all three time periods. Most items that are necessary to assess the methodologic quality of a study were reported by fewer than 20% of the studies. Improvements over time were seen for some items, including the description of statistics methods used, reporting of primary research outcomes, performance of power calculations, method of randomization and concealment allocation, and having performed intention-to-treat analysis. Conclusions: Despite recent improvements, reporting levels of CONSORT items in RCTs involving patients with Hodgkin lymphoma remain unsatisfactory. Further concerted action by journal editors, learned societies, and medical schools is necessary to make authors even more aware of the need to improve the reporting RCTs in medical journals to allow assessment of validity of published clinical research. [J Natl Cancer Inst 2006;98:620–5]

The dissemination of biomedical information and publication of research results is integral to scientific endeavor and is closely linked with the historical development of clinical therapeutic trials, beginning with the recording of medical interventions used by the ancient Egyptians (e.g., bandaging and stitching of wounds circa 1600 bc (1)). One important way to convey and develop scientific knowledge is through empirical studies that compare the effectiveness of behavioral, biologic, and chemical interventions in treating certain medical conditions.

Within the hierarchy of clinical studies, the randomized controlled trial (RCT) is considered the gold standard for establishing effectiveness because it minimizes bias in evaluating new treatment strategies (2–4). RCTs represent a key research activity with the potential to improve the quality of health care and control costs through careful comparison of alternative treatments (5–7). However, the flood of information available in biomedical journals during the past 50 years has been enormous and creates problems in a variety of areas, e.g., publication and selection bias and retraction of invalid literature (8,9). Reporting standards provide a way to bring consistency to the literature and reduce problems that can arise from differences in language conventions that depend on type of study and specialty area (10).

In 1982, DerSimonian et al. (11) suggested that the reporting of clinical trials could be greatly improved by providing authors with a list of items that are expected to be reported. A group of journal editors, clinical trialists, and methodologists subsequently published a common checklist for items to include in reports of RCTs in 1996, known as the Consolidated Standards of Reporting Trials (CONSORT) statement. Work on a revised checklist started in 1999 that, on completion in 2001 (12), featured 22 items to include when reporting an RCT. Its use is recommended by the International Committee of Medical Journal Editors, the Council of Science Editors, and the World Association of Medical Editors (13); to date, more than 150 biomedical journals, including the Journal of the National Cancer Institute, have adopted these recommendations (14).

A number of publications have studied the quality of reports of RCTs in subspecialties of medicine (15–30). However, no study, to our knowledge, has investigated RCTs focusing on Hodgkin disease alone and analyzed changes in reporting over multiple decades. Furthermore, empirical evidence on the quality assessment of RCTs in the field of medical oncology is scanty. In this article, we report the results of our examination of the quality of reporting of RCTs involving patients with Hodgkin lymphoma in the context of the CONSORT statement. We chose this cancer because it is a complex clinical condition that requires sophisticated multimodal interventions and treatment schedules. Consequently, clinical trials in Hodgkin lymphoma are often conducted by cooperative groups, a setting that is presumed to be high in quality of design, conduct, and evaluation of coordinated trials. In addition, trials in Hodgkin lymphoma have been carried out for many years, providing the opportunity to study changes in reporting over a relatively long time.

Methods

Search Strategies and Study Identification

Online searches of the Medline, EMBASE, Cochrane Central Register of Controlled Trials (CENTRAL), Institute of Scientific Information Web of Science (ISI WoS), PsychIndex, Allied and Complementary Medicine (AMed), and CancerLit
biomedical databases for reports on RCTs involving patients with Hodgkin lymphoma were performed through May 2003, supplemented by hand searching of reference lists. To ensure that we would find as many trials as possible we searched text and index words with the truncated version “hodgkin*.” Results were combined (“AND”) with the Cochrane Collaboration sensitive search strategy for RCTs (31). References were screened independently by two persons (S. Trelle and T. Kober) for eligibility without language restriction. Trials were eligible if they had randomly assigned participants prospectively to at least two treatment arms and included patients with Hodgkin lymphoma. Trials were classified as reports if they were published as full or short papers, editorials, or letters in a regular issue or supplement of a biomedical journal or book from 1966 through 2002.

A total of 20499 potentially eligible references were identified (Fig. 1), of which 10397 were found to be duplicates. The remaining 10102 unique citations were screened independently by two persons (S. Trelle and T. Kober). Inter-rater agreement was assessed using the kappa statistic on randomly selected references on 16 occasions during the screening process and was classified as “good” (median for kappa: 0.6).

After eligibility screening, 8809 citations did not fit the inclusion criteria (e.g., they were RCTs in diseases or conditions other than Hodgkin lymphoma; nonrandomized trials, narrative reviews, or meta-analyses in Hodgkin lymphoma; RCTs in non-Hodgkin lymphoma; or not relevant to the topic), leaving 1293 citations requiring complete full-text evaluation. Of those, nine (0.7%) could not be retrieved. The remaining 1284 references were obtained in hard copy and screened further. Of these, 814 articles were excluded because they also did not fulfill the inclusion criteria, leaving 470 citations for additional analysis. Of these, 27 (5.7%) that were not in English or German were omitted, as was one that turned out to be a duplicate. In addition, 200 abstracts were not considered further due to their limited information value (32). Consequently, a total of 242 reports remained for analysis.
Table 1. Frequency of reporting of 14 data items in a total of 242 randomized clinical trials in Hodgkin lymphoma by publication period (pre- and post-CONSORT and combined)*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Power calculation</td>
<td>30/225; 13% (9% to 18%)</td>
<td>1/81 (1%)</td>
<td>9/69 (13%)</td>
<td>10/85 (12%)</td>
<td>.030</td>
</tr>
<tr>
<td>Intention to treat</td>
<td>28/241; 12% (8% to 16%)</td>
<td>0/83 (0%)</td>
<td>7/70 (10%)</td>
<td>21/88 (24%)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Primary outcomes</td>
<td>105/236; 44% (38% to 51%)</td>
<td>12/80 (15%)</td>
<td>32/69 (46%)</td>
<td>61/87 (70%)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Precision of estimated effect size</td>
<td>52/240; 22% (17% to 27%)</td>
<td>4/82 (5%)</td>
<td>9/69 (13%)</td>
<td>39/89 (44%)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Statistical methods</td>
<td>199/240; 83% (78% to 87%)</td>
<td>63/83 (76%)</td>
<td>55/70 (79%)</td>
<td>81/87 (93%)</td>
<td>.003</td>
</tr>
<tr>
<td>Randomized in title/abstract</td>
<td>211/242; 87% (82% to 91%)</td>
<td>65/83 (78%)</td>
<td>63/70 (90%)</td>
<td>83/89 (93%)</td>
<td>.004</td>
</tr>
<tr>
<td>Conflict of interest</td>
<td>10/240; 4% (2% to 8%)</td>
<td>0/83 (0%)</td>
<td>4/69 (6%)</td>
<td>6/88 (7%)</td>
<td>.027</td>
</tr>
<tr>
<td>Method of randomization</td>
<td>31/226; 14% (10% to 19%)</td>
<td>7/79 (9%)</td>
<td>8/66 (12%)</td>
<td>16/81 (20%)</td>
<td>.045</td>
</tr>
<tr>
<td>Concealment of allocation</td>
<td>30/234; 13% (9% to 18%)</td>
<td>5/82 (6%)</td>
<td>11/66 (17%)</td>
<td>14/86 (16%)</td>
<td>.059</td>
</tr>
<tr>
<td>Withdrawals and dropouts</td>
<td>134/230; 58% (52% to 65%)</td>
<td>42/77 (55%)</td>
<td>37/67 (55%)</td>
<td>55/88 (63%)</td>
<td>.294</td>
</tr>
<tr>
<td>Scientific background and purpose</td>
<td>226/235; 96% (93% to 98%)</td>
<td>75/79 (95%)</td>
<td>68/69 (99%)</td>
<td>83/87 (95%)</td>
<td>.901</td>
</tr>
<tr>
<td>Summary results</td>
<td>238/240; 99% (97% to 100%)</td>
<td>81/82 (99%)</td>
<td>68/69 (99%)</td>
<td>89/89 (100%)</td>
<td>.373</td>
</tr>
<tr>
<td>Adverse events</td>
<td>187/227; 82% (77% to 87%)</td>
<td>66/79 (84%)</td>
<td>51/64 (80%)</td>
<td>70/84 (83%)</td>
<td>.981</td>
</tr>
<tr>
<td>Details of intervention in each arm</td>
<td>231/241; 96% (93% to 98%)</td>
<td>80/83 (96%)</td>
<td>67/70 (96%)</td>
<td>84/88 (95%)</td>
<td>.761</td>
</tr>
</tbody>
</table>

*CI = confidence interval. CONSORT = Consolidated Standards of Reporting Trials.
†The number of articles reporting the item/total number of articles, percentage, and 95% confidence intervals. Some data are missing for most items (median number of reports for which data are missing = 4; range = 0–16).
‡P values were obtained from chi-square tests of associations between proportions for reporting an item and publication period across the three periods.

Reporting Assessment Tool

A data extraction sheet based on CONSORT reporting items was developed. Although all items in the CONSORT checklist are considered important to help to improve the quality of reports of RCTs, some are more subjective than others (e.g., assessment of whether Discussion sections of manuscripts addressed interpretability of results). We did not consider items that are mainly relevant for assessing external validity of a study (e.g., eligibility criteria or baseline characteristics) because our aim was to survey the reporting level of items that are important to assess potential biases. We therefore chose to analyze 13 of the 22 items of the CONSORT statement (Table 1). We also analyzed reporting of an additional item that is not included in the CONSORT statement, namely, conflict of interest. This item was chosen for its relevance in terms to reflect general transparency in scientific writing and was adopted from The Lancet as an acceptable reporting item (33).

Items were investigated in terms of whether they were reported, not whether they were actually carried out during the trial. Response alternatives to each question were: yes, no, unclear, and not applicable (NA). Unclear and NA responses were both coded as missing data.

Statistical Methods

Reporting items to be evaluated were derived from the data extraction sheet and analyzed with Stata (Version 9.1) by descriptive summary statistics and exact confidence intervals. To test for reporting differences over time, reports were grouped in three publication periods, i.e., 1966–1988 (pre-CONSORT 1), 1989–1995 (pre-CONSORT 2), and 1996–2002 (post-CONSORT). The cut points for time periods were chosen to reflect the low number of publications in the mid-1960s to mid-1970s. The average annual publication rate of RCTs in Hodgkin lymphoma increased from 1.25 publications per year for the period 1966–1974 to 14.6 for the period 1996–2002, representing a mean increase of approximately eight publications each year over the entire period under investigation. Because discussions about reporting quality of trials increased in the 1980s (11,12), we used a second pre-CONSORT period to investigate possible changes that occurred before the introduction of CONSORT.

Comparisons among periods were made using a chi-square test for trend. Clustering of articles in a journal or by study group was not taken into account in the analyses. The cutoff point for statistical significance was set at the two-sided .05 level. No adjustments for multiple comparisons were used because of the exploratory nature of the hypotheses tests.

Results

Of the 242 study reports that we analyzed, 83 were published in 1966–1988 (pre-CONSORT 1), 70 in 1989–1995 (pre-CONSORT 2, and 89 in 1996–2002 (post-CONSORT). Table 1 shows the frequency of reporting of the 14 items in our data extraction sheet for each of these three periods and for the combined period.

Only six items were reported by 75% or more of the studies in all of the time periods (Table 1). These include reporting of randomization in the title or abstract, descriptions of statistical methods, description of the scientific background and purpose of the study, reporting summary results, reporting adverse events, and reporting the details of the intended intervention in each group.

By contrast, a number of items identified by data extraction were reported by only a small percentage of the trials in all three periods. For example, only 14% (31 of 226) of reports provided a description of the randomization process, although there was some improvement after CONSORT was implemented. Similarly, only 30 of 234 articles (13%) provided details about whether and how concealment of allocation was carried out. A trend test across all three periods showed better reporting of allocation concealment after CONSORT. Only 13% (30 of 235) of all studies provided a statement or explanation of how the study power was calculated. However, reporting of this information improved in the post-CONSORT period. Similarly, although only a small number of studies (28 of 241, or 12%) stated the performed data analysis explicitly as intention-to-treat, reporting of this item also showed a statistically significant improvement after CONSORT.
Only 4% of articles (10 of 240) provided a conflict of interest statement. Although there was some improvement over time, it was only moderate. Two additional items showed increases in reporting frequency after CONSORT. These were explicit reporting of primary outcomes and estimates of the precision of estimated effect size (i.e., presentation of 95% confidence intervals).

We also performed a stratified analysis according to whether journals have or have not adopted the CONSORT statement (data not shown). Overall, improvements over time were mostly seen in journals not adopting the CONSORT statement. However, the number of articles in journals that promote CONSORT was low, and reporting quality was already better in the pre-CONSORT era in these journals than in journals that did not go on to adopt CONSORT. A comparison between journals adopting CONSORT and journals not adopting these recommendations for the post-CONSORT era only showed no difference in reporting quality for any of the assessed items (data not shown).

Reporting of results according to the intention-to-treat principle was analyzed in more detail because deviations from this principle can lead to overoptimistic and biased results. We therefore examined closely the subset of 29 studies that explicitly stated that analyses were carried out according to this principle. Of these studies, only 17 (59%) appeared to actually analyze participants as randomly assigned.

**DISCUSSION**

Our research shows that essential methodologic aspects of RCTs are seldom described in published reports, making it impossible for the reader to assess their validity. We found statistically significant improvements over time in some reporting items but several important methodologic descriptions (of allocation concealment and of the randomization method) improved only minimally. Nevertheless, a consistent trend to better reporting was observed in all methodologic aspects of study reports. This trend started even before the publication of CONSORT and is probably related to the increasing discussion about methodologic quality of clinical research in the 1980s. The CONSORT statement may be viewed as a result of these discussions and improvements even before its publication. The observed reporting level seen in this study may therefore be more related to the general scientific discussion and less to the publication of CONSORT itself. Although the direct impact of CONSORT remains undetermined from this study, it seems clear that CONSORT provides an explicit and therefore useful framework for authors and editors. Its publication and the preceding discussions may have increased awareness on the side of authors and editors, resulting in improved reporting.

Data on the quality of reports of RCTs in subspecialties of medicine are increasingly available. Although surveys and studies similar to this one have been conducted previously, this research is the first, to our knowledge, to investigate the reporting of RCTs with particular reference to CONSORT for Hodgkin lymphoma only, covering a period of almost 40 years. One of the first studies to investigate the extent of ambiguous and unclear reporting in the medical literature was research undertaken by Mulrow et al. (15), which assessed 50 papers published from June 1985 to June 1986 in four leading journals based on eight explicit information criteria. Of the 50 publications, 17 satisfied three of the eight criteria, 32 satisfied four or five, and one satisfied six. Moher et al. (27) were among the first to investigate whether the use of the CONSORT statement was associated with an improvement in the quality of reports of RCTs. They compared the quality, as assessed by the number of CONSORT items included in the report, of 71 RCTs published in BMJ, JAMA, The Lancet, and The New England Journal of Medicine in 1994 (pre-CONSORT) with that of 77 RCTs published in the same journals in 1998 (post-CONSORT). For three of the four journals, statistically significantly more CONSORT checklist items were included in RCT reports in 1998 than in 1994 (mean number pre-CONSORT = 23.4; mean change = 3.7). Devereaux et al. (26) reported an evaluation of 105 RCT reports published in 29 biomedical journals. They analyzed the quality of reporting of 11 key methodologic factors in relation to whether a journal had adopted CONSORT. The analysis suggested that articles in journals adhering to CONSORT guidelines included a statistically significantly higher number of reporting factors than those in nonpromoter journals ($P = .03$).

A more recent study, by Mills et al. (14), investigated the extent to which four clinical pharmacology journals had implemented specific CONSORT recommendations. Among 193 RCTs, the following items were reported in the majority of studies: use of intention-to-treat analysis, 79%; description of withdrawals, 92%; description of adverse events, 71%; and sources of funding, 56%. By contrast, several items were poorly reported: description of method of randomization, 17%, and allocation concealment, 3%. In an evaluation of reports of RCTs in myeloma (25), only 10 of 136 (7%) were analyzed according to intention-to-treat principles, 12 of 136 (9%) reported a power analysis, 35 of 111 (32%) reported adequately concealed treatment allocation, and 106 of 136 (78%) provided a detailed description of patient withdrawals.

Our study has several limitations. One is that we did not include all 22 reporting items listed in the CONSORT checklist. However, to our knowledge our study included more items than any previous study; in addition, we investigated one aspect not included in CONSORT, that is, the declaration of any financial or personal conflict of interests. The results in reporting conflict of interest post-CONSORT are, in our view, most likely influenced by the *Uniform Requirements for Manuscripts Submitted to Biomedical Journals*, a document developed by the International Committee of Medical Editors (34). Another limitation of our study was that it was designed only to evaluate the reporting quality of RCTs overall and not to assess the quality of the individual study design or to assess how study design affects intervention outcomes (23,35,36). As empirical research has shown, the quality of the trial may well be different from the quality of the study report (37). For example, Soares et al. (24) evaluated the quality of 58 published reports of RCTs completed by the Radiation Therapy Oncology Group. This analysis showed that only 9% of published reports provided information on sample size calculation, although this information was available in 44% of the group’s protocols. Yet another potential limitation is that we assessed only publications in English and German, which may contribute to introduce publication bias. However, only 10% of the eligible articles were reports in other languages, and it is unlikely that their inclusion would have changed the overall results. Another limitation would arise if Hodgkin lymphoma is not representative for reporting RCTs in other malignant diseases. However, because investigations about reporting quality in other medical conditions have reached conclusions similar to ours, we...
believe that RCTs in Hodgkin lymphoma may well be representative of trials in cancer generally.

In summary, our findings indicate that reports of RCTs involving patients with Hodgkin lymphoma published after 1996 do not conform with CONSORT recommendations. Consequently, users of such trial reports may still find it difficult to easily identify, assess, and synthesize all data required to make a comprehensive and thorough judgment about the benefits and harms of the interventions tested. Moreover, the application of the results in clinical practice is hampered if not prevented, because their validity is not assessable.

Although we did not investigate the relationship between CONSORT adherence and study quality or the validity of study results since 1996, our study identified an undesirable level of ambiguity and uncertainty in the reporting of RCTs. This is particularly evident in the number of missing responses for most data items under investigation, ranging from 1% to 7% (Table 1). It is imperative that clinicians be able to maintain up-to-date information about methods of data presentation, such as CONSORT, that are applicable to RCTs. The onus is on both parties—those who submit manuscripts and those who read them during the peer review and publication stages. Because reporting quality is still low after the publication of CONSORT, editors, in particular, should be encouraged to scrutinize manuscripts more carefully and to remind and guide authors in the use of CONSORT as a prerequisite for publication. To foster better reporting, all journal recommendations for preferred presentation and analysis of data should be described in the Information for Contributors or Authors. Whenever possible, recommendations should be based on evidence about methods of data presentation, such as CONSORT, that are readable and most likely to be interpreted correctly by readers. Editors should keep themselves informed of this research and adapt their recommendations as it evolves. Empirical evidence shows that there is room for improvement on the side of the journals (38).

REFERENCES

Conflict of interest statements for authors. Lancet 2004;363:2–3.


NOTES

The Cochrane Haematological Malignancies Group is part of the Competence Network Malignant Lymphomas and is funded by the German Ministry of Education and Research (BMBF).

Manuscript received August 16, 2005; revised February 2, 2006; accepted March 22, 2006.