Reporting of Randomized Clinical Trial Descriptors and Use of Structured Abstracts

Roberta W. Scherer, PhD; Barbara Crawley, MS

Context.—Structured abstracts, that is, abstracts that describe a study using requisite content headings, provide more informative content. Concomitant reporting in the text of the report might improve with structured abstract use because of increased awareness by authors or editors of important study areas associated with content headings.

Objective.—To assess whether structured abstract use is associated with improved reporting of randomized clinical trials.


Main Outcome Measures.—We measured the inclusion of 56 criteria derived from Consolidated Standards of Reporting Trials (CONSORT) descriptors (JAMA 1996;276:637-639) in the text of each report and calculated the number of criteria included per report and the proportion of reports including individual criteria. Reports with structured abstracts were compared with those without, and reports published in 1993 and 1994 in the American Journal of Ophthalmology were compared with those published in 1991 and 1992.

Results.—The mean (SEM) number of criteria included by authors was 15.8 (0.4) per report in 125 trial reports. We found no difference in the mean number of criteria included or the proportion of reports that included specific criteria by journal. Following structured abstract use, there was no difference in either the mean number of criteria per report or the proportion of reports including a majority of criteria within each CONSORT subheading. Four criteria were included more often and 2 less often following structured abstract use in individual journals.

Conclusion.—Using CONSORT descriptor criteria to evaluate reporting quality, we found no difference in text reporting associated with structured abstract use in the journals examined.

From the Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, Baltimore.

Presented at the Third International Congress on Peer Review in Biomedical Publication, Prague, Czech Republic, September 20, 1997.

Reprints: Roberta W. Scherer, PhD, Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, Century Bldg, First Floor, 506 W Fayette St, Baltimore, MD 21201 (e-mail: rscherer@epi.ummARYLAND.edu).

the possibility that “[abstract] structure reminds authors . . . of the necessity of providing each category of information.” By focusing on the content headings required by a structured abstract, authors or editors might incorporate descriptors more consistently in the manuscript text. Our objective was to assess whether the use of structured abstracts is associated with an improvement in the overall reporting of ophthalmology randomized controlled trials (RCTs).

**METHODS**

**Materials**

We selected 2 US ophthalmology journals that we previously found to regularly report large numbers of RCTs and that recently revised their Instructions to Authors to require structured abstracts, the *Archives of Ophthalmology* and *Ophthalmology*. We manually searched for reports describing RCTs the year preceding, of, and following the first appearance of structured abstracts (1992, 1993, and 1994, respectively, in the *Archives of Ophthalmology* and 1991, 1992, and 1993, respectively, in *Ophthalmology*). To monitor changes over time, we searched the *American Journal of Ophthalmology* (a US ophthalmology journal that publishes similar numbers of RCTs) and that did not require structured abstracts until late 1994 for RCT reports from 1991 through 1994.

Hand searching was carried out by 2 independent, trained readers who examined each full-length report for RCT status in the selected journals for the specified years. We defined RCT as a controlled experiment designed to evaluate an intervention or diagnostic tool, using...
a random method to assign individuals, eyes, or some other unit to a test or comparison group. We included quasi-randomized clinical trials, ie, those employing a method of assignment (eg, alternation) designed to avoid bias. Reports of RCTs that did not include data by randomized treatment group (eg, validation of a method used for measuring an outcome) were excluded.

We found 154 reports of RCTs. Five papers were excluded; 1 had no abstract, 2 presented data on subsets of patients, and 2 examined methods to measure outcomes. We also excluded 24 reports because the authors stated that an abbreviated methodological description was provided since the methods had previously been described.

### Extraction of RCT Design and Operational Characteristics

The Consolidated Standards of Reporting Trials (CONSORT) statement describes RCT descriptors and provides a flowchart showing patient entry and follow-up that are recommended for inclusion in every RCT report. Using CONSORT descriptors as a “gold standard” to evaluate reporting quality in each article, we scored the presence of 56 criteria (each flowchart block or descriptor, Table 1) in individual reports as yes, no, or not applicable. We selected 9 criteria (with daggers in Table 1) corresponding to common abstract content headings to measure inclusion in each abstract.

Because some study- or journal-specific characteristics could influence the inclusion of CONSORT statement descriptors, we extracted information by journal about study and report characteristics, including purpose of intervention, multicenter status, type of test intervention, sample size, group or individual authorship, length of report (number of pages), and length of methods section (number of pages).

### Analyses

Data were entered into a database (Paradox Version 4.0, Borland International, Scotts Valley, Calif), and exported to a statistical program (SAS Version 6.2, SAS Institute Inc, Cary, NC). We calculated the number of criteria included in each report, the proportion of reports that included specific criteria, and the number of reports that included more than the majority of criteria within each CONSORT subheading (introduction, protocol, etc). We compared structured with unstructured abstracts, and reports published in 1991 and 1992 with those published in 1993 and 1994 (American Journal of Ophthalmology), using the Student t test or χ² tests. Odds ratios (ORs) with 95% confidence intervals (CIs) are presented.

### RESULTS

#### Description of RCTs by Journal

We found that RCTs reported in the Archives of Ophthalmology, compared with Ophthalmology and the American Journal of Ophthalmology, were more often multicentered (10 [38%] of 26 vs 12 [24%] of 51 and 13 [27%] of 48, respectively; P = .07), evaluated surgical or laser trials more often (7 [27%] of 26 vs 6 [12%] of 51 and 3 [6%] of 48, respectively; P = .001), more frequently devoted more than a single page to the methods section (10 [38%] of 26 vs 9 [18%] of 51 and 12 [25%] of 48, respectively; P = .005), and published more reports with study group authorship (5 [19%] of 26 vs 2 [4%] of 51 and 3 [6%] of 48, respectively; P = .001). Thus, we examined all results separately by journal.

#### Reporting of Descriptors in Text by Journal

Reporting of CONSORT criteria in the text was unimpressive. The mean (SEM) number of criteria included was 15.8 (0.4) of a possible 56; there was little difference among journals (Table 1). Journals were also remarkably similar in the proportion of reports that included specific criteria (Figure 1). Criteria reported in a low proportion of reports in all 3 journals were often associated with CONSORT subheadings associated with RCT methods, such as assignment and masking.

### Comparison by Structured Abstract Use or Over Time

We found no difference in the mean number of criteria that were included in reports with structured abstracts compared with those without (Table 2). Including a majority of criteria within a single CONSORT subheading was positively associated with structured abstracts for “protocol” in the Archives of Ophthalmology (OR, 2.16; 95% CI, 1.12-4.16), negatively associated for “introduction” in Ophthalmology (OR, 0.57; 95% CI, 0.33-0.95), and not associated with later year of publication. Thus, we found no evidence for improvement in inclusion of criteria associated with structured abstracts.

We then calculated the proportion of reports in which a specific criterion was included to see if there were improved reporting by criterion associated with structured abstract use or year of publication. Individual criteria were included more or less often following structured abstract use or with later publication years, but by individual journal (Table 3).

#### Reporting of Descriptors in Abstract of Report

Of the 9 criteria used to evaluate abstract reporting, a mean of 5.0 (0.2) were included in all abstracts from all journals. Structured abstracts were more often associated with inclusion of criteria in the Archives of Ophthalmology (Table 2).

Specific criteria included infrequently in structured abstracts for the Archives of Ophthalmology...
of Ophthalmology, Ophthalmology, and American Journal of Ophthalmology were description study population (6 [67%] of 9, 18 [57%] of 28, and 16 [64%] of 25, respectively); primary outcome (6 [67%] of 9, 8 [29%] of 28, and 12 [48%] of 25, respectively); and number of patients followed up (2 [22%] of 9, 5 [18%] of 28, and 3 [12%] of 25, respectively).

COMMENT

Our results do not support an association of improved text reporting with structured abstracts or later publication in the journals examined. These 3 journals tended to be more alike than different in overall reporting of individual criteria, and there was no consistent pattern of change in reporting associated with structured abstract use or later publication. Possibly, the time period we examined was transitional or represented a lag time while editors or authors were incorporating use of structured abstracts. Haynes et al reported that initial use of structured abstracts found some authors writing abstracts concurrently with manuscript preparation, whereas others did so only at submission or on editor request. Also, no special emphasis was placed on use of structured abstracts for RCTs initially, even with recognized importance for trial reports. Finally, it is possible that our sample size was insufficient to detect subtle changes in reporting, as our results are based on 125 reports from 3 ophthalmology journals; they may not be generalizable to other journals or areas of medicine.

Although we did not detect improvement in overall text reporting, we thought there might be individual criteria reported more frequently when structured abstracts were used, but found no consistent reporting pattern. For example, use of random or trial in the title and rationale for statistical tests were reported more often, but in separate journals. Some criteria were reported less frequently with structured abstracts, and perhaps were viewed as less important when space constraints limited text length. However, any changes we report in inclusion of criteria associated with structured abstract use or later year of publication may be due to chance, given the number of observations.

Although checklists for assessing reporting quality of RCTs were available,15-15 we chose to use CONSORT descriptors as a "gold standard" since it comprises a comprehensive list of criteria. We did not intend to evaluate trial quality. It has been argued, however, that "a well-designed but poorly reported trial could be judged as having low quality,"55 so assessing reporting is an important first step in assessing trial quality.

Our initial search yielded 24 RCT reports that included abbreviated methodological descriptions because methods had previously been reported. Editors are faced with a tension between space limitations and inclusion of all CONSORT descriptors in subsequent RCT reports. Since readers may not have previous reports available, we believe each RCT report should include all CONSORT descriptors to allow independent report evaluation.

Finally, we found a significant improvement in abstract reporting quality when structured abstracts were used in the Archives of Ophthalmology, and some improvement in Ophthalmology. Consistent with findings of others,56-58 we found abstract reporting deficiencies with authors frequently omitting a description of study population, primary outcome, or number of patients followed up.

In summary, we found no improvement in text reporting when structured abstracts were used. Nevertheless, structured abstract use should not be abandoned since abstract report itself is improved using this format. Rather, our results highlight the need for a standard such as the CONSORT statement to enhance RCT text reporting.

We wish to thank Patricia Langenberg, PhD, and Kay Dickersin, PhD, for many helpful suggestions and discussions.

References


Table 3.—Criteria Included in a Significantly Larger or Smaller Proportion of Reports Associated With Structured Abstracts or Later Year of Publication.
We agree with Berman that there may often be honest differences of opinion among informed experts about the value of new treatments. As stated in our article, our study cannot establish whether current practices are right or wrong. However, the wide gap between the recommendations of physicians and the approvals of insurers in the cases under study underscores a potentially far-reaching problem in the US health care system. We believe that insurers often make very careful and fair analyses of information in arriving at coverage decisions. Nevertheless, greater dialogue between insurers and physicians is needed to minimize discrepancies in coverage and to provide optimal patient care.

Dr Yaes asks about consensus among physicians for the cases studied and about the value of a treatment such as GH, which is used for a non–life-threatening condition. We focused on GH therapy as a key example of emerging treatments that are semielective, relate at least in part to quality of life, and for which consensus about optimal utilization is lacking. As we pointed out in our article, many emerging therapies share similar characteristics, (eg, treatments for infertility, impotence, and aging). For GH therapy, lack of consensus exists for severe cases of idiopathic short stature, whereas there is much more consensus among physicians for GH treatment of children with Turner syndrome and renal failure. Nevertheless, we found significant discrepancies between physician recommendations and insurer coverage for Turner syndrome and renal failure, as for idiopathic short stature.

Ms Tesch and Dr Yaes differ widely in their viewpoints and underscore the fact that the value of treatments for extreme short stature (or other conditions that are semielective and address, to some extent, quality of life) is based, in part, on the degree to which the underlying condition is perceived as representing a form of morbidity. As more such treatments emerge, clinicians will increasingly be confronted with difficult questions about their use. Our study illustrates the real-life difficulties in delivering optimal and equitable health care for nonemergency conditions. We believe that the discrepancy between physician treatment recommendations and insurance coverage constitutes an important challenge to health care delivery and that patients deserve increased and constructive dialogue between physicians and insurers in arriving at coverage decisions.

Leona Cuttler, MD
Beth S. Finkelstein, PhD
J. B. Silvers, PhD
Duncan Neuhauser, PhD
Ursula Marrero, MSSA
Case Western Reserve University
Cleveland, Ohio

In Reply.—We agree with Dr Berman that insurers vary substantially in the procedures they use to develop and implement policy regarding coverage for new technologies (and other treatments) and in the values they apply in making their decisions. We also agree that many insurers make serious efforts to address the 3 conditions we believe are necessary to create a public sense of legitimacy and fairness for limit-setting policies: (1) articulate the rationale behind their policies, (2) make the rationale public, and (3) provide opportunities for clinicians and members to appeal policies and specific decisions. However, even in the outstanding programs we were privileged to study in our research on policymaking, we and the program leaders agreed there was room for significant improvement.

Ms Tesch’s letter on behalf of the Turner’s Syndrome Society provides an example of well-conceived advocacy. The fact that GH has been recognized by the Food and Drug Administration as effective in the treatment of Turner syndrome puts a burden of explanation onto an insurer who chooses not to cover it. In asking whether decisions to deny coverage reflect “reasonable cost-benefit analysis or an inappropriate denial of care,” Tesch raises the key policy question in useful terms. In a poor country, denial would be easily defendable in terms of competing priorities, but in a country as wealthy as the United States, the case for denying coverage is more difficult to justify.

Dr Yaes notes the degree of controversy that surrounds use of GH to treat idiopathic short stature. We disagree with his view that the controversy makes GH a bad focus of study. The controversy actually makes GH an excellent “biopsy” of policymaking and underscores the importance of articulating the rationale for noncoverage. If denial of coverage for GH in idiopathic short stature is based on doubt about its efficacy, proof of efficacy would refute the objection. If the denial is based on fear of adverse effects, an informed patient and family might argue that they understand and choose to accept the risks. If the denial is based on regarding any benefits as “purely cosmetic,” then the issue is one of our understanding of short stature and the goals of medicine. Yaes’ comments are a contribution to the kind of deliberative process on which coverage policies should be based.

James E. Sabin, MD
Harvard Pilgrim Health Care
and Harvard Medical School
Norman Daniels, PhD
Tufts University
Boston, Mass


CORRECTIONS

Error in Reference List.—In the article entitled “Reporting of Randomized Clinical Trial Descriptors and Use of Structured Abstracts,” published in the July 15, 1998, Peer Review theme issue of THE JOURNAL (1998;280:269-272), an error was made in the reference list. On page 272, the second reference 9 should be deleted.

Incorrect Unit of Measure.—In the Letter by Rimm et al entitled “Relationship of Dietary Folate and Vitamin B6 With Coronary Heart Disease in Women” published in the August 5, 1998, issue of THE JOURNAL (1998;280:418-419), there was an incorrect unit of measure listed. On page 419, the last sentence of the second full paragraph reads, “The concern raised by Herbert about masking anemia due to vitamin B12 deficiency does not apply to the usual amount of folate (400 µg/d) consumed by women in our cohort as part of a multiple vitamin.” The amount of folate should have been 400 µg/d.