Quality Evaluation of Controlled Clinical Information Service Trials

E. Andrew Balas, M.D., Ph.D., Suzanne M. Austin, M.H.A.,
Gordon D. Brown, Ph.D., Joyce A. Mitchell, Ph.D.

Medical Informatics Group and Health Services Management Program,
University of Missouri-Columbia

Randomized controlled clinical trials are increasingly accepted as tools of computer technology assessment and, therefore, quality evaluation of trials has great theoretical and practical significance. The purpose of this study was to assist the design of evaluation studies and synthesis of published results by developing and validating an easy-to-use quality scoring method. The development of the new scoring system was based on the available quality evaluation methods and the analysis of 19 trial reports registered in the Columbia Registry of Controlled Clinical Information Service Trials. First critical aspects and afterwards the levels of quality were defined. In spite of the fact that all quality requirements were met by some trials, the average overall quality score was 52.6 (±8.7) per cent. The minimum score was 37 and the maximum was 72 per cent. Data collection and site/sample definition were better in the good quality trials, but improvement in statistical analysis was erratic. The quality scoring method was validated by using another sample of 20 registered trials. While the number of published controlled clinical trials is increasing in medical informatics, the analysis was unable to demonstrate a significant positive correlation between the quality and year of publication.

INTRODUCTION

Clinical trials can provide the most reliable information about the practical value of computer-assisted information systems [1], [2], [3]. Randomized controlled clinical trials represent the planned experimental approach in clinical medicine and can avoid the dangers of selection bias, nonstandard definitions, missing data, or multiple comparisons [4]. The number of published reports on randomized controlled clinical trials of computer applications is rapidly increasing [3], [5], [6].

Quality of a clinical trial is often interpreted as the repeatability and generalizability of the research results. Fortunately, these aspects can be measured and summarized in a single quality indicator by using an appropriate evaluation form (e.g., [7]). Such quality evaluation forms are useful checklists for designers of controlled clinical trials, readers of trial reports, and researchers who synthesize published results of clinical trials (expert reviews, meta-analyses). Synthesis of research results requires quality filtering of studies, especially when unpublished trials are also included.

In spite of the fact that controlled health services research trials, like computer trials, apply several techniques developed for clinical drug evaluation studies, there are important differences.which have to be considered in the evaluation of quality. Certainly, these differences limit the use of available trial quality evaluation methods. For example, compliance is rarely an issue of trial quality but a commonly used effect variable (end point) in computer technology assessment (e.g., compliance of physicians with preventive care guidelines). Selection of biological equivalent, taste of placebo, or blinding of physician regarding the intervention do not play a role in controlled health services research trials. On the other hand, specific description of the trial site is a critical piece of information in health services research. Considering the danger of subjective bias and inter-rater variability, statements like "fair", "poor", or "inappropriate" are not

0195-4210/92/$5.00

specific enough to evaluate quality aspects. New requirements of the increasingly popular meta-analyses also urge a revision which streamlines score calculation and makes the quality evaluation faster and more effective. The significant differences between drug trials and information service trials limit the use of available scoring systems and urge a major revision.

The aim of this study was to develop and validate a clinical trial quality scoring system which can meet the needs of the evaluation of computer trials. Such scoring could be used to evaluate the quality of published trial reports and provide quick information about the validity of recommendations.

METHODS

The development of quality scoring methods was based on the analyses of 39 trial reports registered in the Columbia Registry [3], the earlier published scoring method of Chalmers et al. [7], and the various requirements of research synthesis, particularly meta-analysis of clinical trials [8].

To provide an easy-to-use tool for evaluation, 29 questions on important quality aspects of controlled health services research trials are compiled in an evaluation form. The following is a list of some of the specific questions which are characteristic to the quality evaluation of information service trials:

- The site of the trial is a critical information in evaluating the generalizability of results from health services research. The differences between managed health care organizations and fee-for-service solo practices, residents and independent practitioners, and the type of medical specialty can substantially alter the feasibility and efficiency of certain interventions in various health care organizations.

- Clinical trials are often based on an inadequate sample and, therefore, the results cannot be used directly. The results of small sample trials may be valuable, but meta-analysis might be needed to increase sample size and make the case-mix more representative. Therefore, the adequacy of sample size has to be evaluated to validate the results.

- Replicable description of the tested intervention and unambiguous specification of effect variables are critical for the dissemination and implementation of the trial results. Unfortunately, vague or incomplete definitions are frequent in trial reports and, therefore, these aspects have to be carefully evaluated.

- Complete reporting of the numeric results is not only an important supportive evidence for the conclusions, but also necessary for further meta-analyses. Therefore, tabular presentation of data has to include numerator and denominator of frequencies, mean and standard deviation of continuous variables.

The total points awarded in the revised questionnaire add up to 100 and the scoring system has been simplified by including specific questions and responses. This has eliminated the need to assign different weights to different categories of questions when all of the questions are not applicable. In addition, several of the questions in the new questionnaire have been assigned different point values than they received in earlier questionnaires. For example, questions relating to the prior estimate of sample size (power calculation), testing randomization, proper retrospective analysis, and the results of prerandomization data analysis receive a greater number of points in the revised questionnaire. On the other hand, questions relating to the blinding of physicians regarding results, the blinding of patients, the handling of withdrawals, and the timing of events including the dates of starting and stopping receive fewer points in the revised questionnaire.

The questions can be categorized as receiving a possible ten points, five points, three points or two points. Only one question could have received a score of ten points. This question referred to the method of randomization used in the study. Twelve question could receive a maximum of five points each. The questions receiving five points are as follows: description of the sites, description of sampling, definition of sample size, description of intervention, description of the main effect variables, description of secondary variables, measurement of effect is blinded to intervention,

	Frequency
Description of the site(s) includes (i) medical specialty, (ii) inpatient or outpatient care, (iii) academic or non-academic center, (iv) reimbursement method, (v) for-profit or not-for-profit organization:	
[5] All listed criteria	6
[3] Two to five criteria	13
[0] Less than two criteria	0
Description of intervention includes (i) persons targeted, (ii) timing and periodicity, (iii) rules and formulas, (iv) replicable description of content	
[5] All listed criteria	18
[3] Two to three criteria	2
[0] Less than two criteria	0
Description of the main effect variables	
[5] Clear definition for each variable and description of methods and results can be matched	17
[3] Vague definition or lack of matching	3
[0] No definition	0
Numeric table of effect variables (mean-standard deviation or numerator-denominator)	
[5] Presented for each effect variable	10
[3] Partially available	1
[0] No or graphics only	9

Table I Distribution of quality scores

testing randomization, numeric table of effect variables, ratio of withdrawals after enrollment,

data on possible adverse effects, and the analysis of the main effect variables. Six questions could receive a maximum of three points each, while the remaining six questions could receive an ideal two points each.

Two samples of controlled clinical trials testing computerized interventions were used in this study. Nineteen trials were included in sample A and 20 trials were included in Sample B. The exploratory phase of the development was based on the analysis of Sample A. Sample B was used to confirm the validity of the new scoring system.

RESULTS

Sample A was the basis of the scoring system development and the following table and figures illustrate the analyses of the trial quality.

Table I summarizes the distribution of answers in Sample A and illustrates the method of developing the scoring system. Four aspects, which are closely related to the practical message of the evaluated trials were selected for this table. Some trials have reached the maximum score in each of the individual aspects. This indicates the expectations used when scoring the trial reports are realistic. However, the cumulative scores of the trial reports remained weak indicating inconsistencies in the methodology.

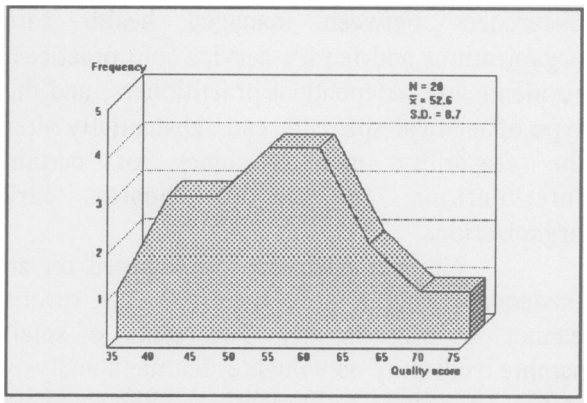

Figure 1. Histogram of scores

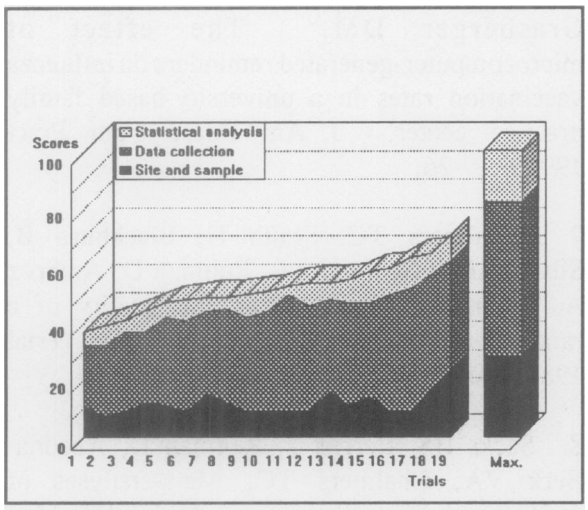

Figure 2. Composition of scores

In Sample A, the histogram of the overall quality scores follows a distribution similar to normal (Fig. 1). The minimum overall score was 37 and the maximum was 72. None of the trials reached the 81-90 per cent ("B") or 91-100 per cent ("A") level which indicates opportunities for improvement. Again, the quality expectations were not unrealistic because the best scores were achieved by one or more trials in nearly all aspect categories. However, consistently high quality across all aspects has not been achieved by any of the evaluated trials.

There are three critical aspect areas of trial methodology: site and sample, data collection, and statistical analysis. In the final phase of the development, the composition of the overall score was evaluated and compared with the ideal scores (Fig. 2). After ranking the evaluated trials according to the overall scores, the contributions to quality improvement can be analyzed. Fig. 2 illustrates that improvement in data collection techniques was the primary contributor of better overall scores. Site and sample definitions, which include randomization, were improved only in the best trials. However, quality of the statistical analysis did not follow the general trend.

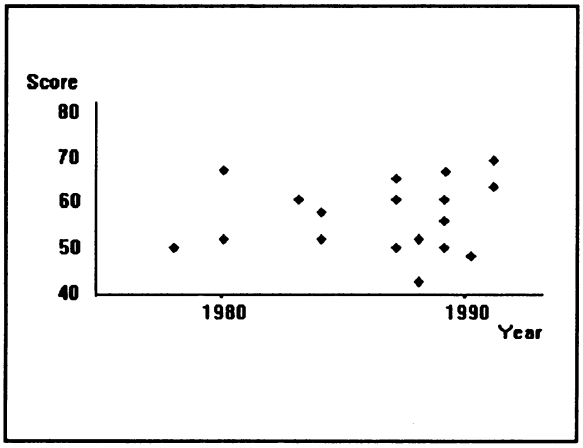

Figure 3. Scores and year of publication

The new quality scoring system was analyzed by using Sample B. The average score on this sample was 58.7 ($\pm$7.7). The minimum score was 43 and the maximum was 71. These results are significantly higher than the scores on Sample A (t test $p<.05$). However, no improvement can be demonstrated in the quality of trials over time (Fig. 3). There was no correlation between the overall score and year of publication ($r=.15$).

DISCUSSION

Our results indicate that the quality of clinical trials testing computer-assisted information interventions needs substantial improvement. Deficiencies of the quality of controlled trials have been repeatedly highlighted by several publications [9], [10], [11], [12], [13]. This study provided more specific information about the common weaknesses of the trials in medical informatics. Particularly, site and sample definitions including randomization, and statistical analysis need improvement in most trials.

In the area of health services research and medical informatics, investigators have received little help for the design of controlled clinical trials. Systematic and validated evaluation forms and checklists have not been available. This study documented the lack of progress in trial quality over time and indicated the need for better support in trial design. Our development efforts resulted in a validated quality evaluation form which is available from the authors upon request.

Certainly, the limitation of quality scoring is that it focuses on a few selected

aspects of clinical research and, most frequently, on a single source of information i.e., the published report. Consequently, quality scoring of trials should not be interpreted as an evaluation of the overall quality of a clinical research project.

Acknowledgement: This research was supported in part by a grant from the Research Board of the University of Missouri.

References

1. Haynes RB, Walker CJ. Computer-aided quality assurance: A critical appraisal. Arch Intern Med 1987;147:1297-1301.

2. Wyatt J, Spiegelhalter D. Field trials of medical decision-aids:potential problems and solutions. Clayton PD (ed) Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care. McGraw-Hill, 1991, Washington D.C. pp 3-7.

3. Balas EA, Mitchell JA, Bopp KD, Brown GD, Ewigman BG. The Columbia Registry of Controlled Clinical Computer Trials. In Frisse ME (Ed): Proceedings of the Symposium on Computer Applications in Medical Care, McGraw-Hill, 1992.

4. Byar DP. Why data bases should not replace randomized clinical trials. Biometrics 1980;36:337-42.

5. Peck CC, Sheiner LB, Martin MM, Combs DT, Melmon KL. Computer-assisted digoxin therapy. N Engl J Med 1973;289:441-6.

6. Chambers CV, Balaban DJ, Carlson BL, Grasberger DM. The effect of microcomputer-generated reminders on influenza vaccination rates in a university-based family practice center. J Am Board Fam Pract 1991;4:19-26.

7. Chalmers TC, Smith H, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. A method for assessing the quality of a randomized control trial. Contr Clin Trial 1981;2:31-49.

8. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. N Engl J Med 1987;316:450-5.

9. Meinert CL, Tonascia S, Higgins K. Content of reports on clinical trials: A critical review. Contr Clin Trials 1984;5:328-47.

10. Altman DG, Dore CJ. Randomization and baseline comparisons in clinical trials. Lancet 1990;335:149-53.

11. DerSimonian R, Charette J, McPeek B, Mosteller F. Reporting on methods of clinical trials. N Engl J Med 1982;306:1332-7.

12. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials: A survey of three medical journals. N Engl J Med 1987;317:426-32.

13. Liberati A., Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. J Clin Oncol 1986;4:942-51.