Quality of nonstructured and structured abstracts of original research articles in the *British Medical Journal*, the *Canadian Medical Association Journal* and the *Journal of the American Medical Association*

Anna Taddio, MSc; Tapas Pain, MSc; Frank F. Fassos, MSc; Heather Boon, BScPhm; A. Lane Ilersich, MSc; Thomas R. Einarson, PhD

Objective: To assess and compare the quality of nonstructured and structured abstracts of original research articles in three medical journals.
Design: Blind, criterion-based observational study.
Sample: Random sample of 300 abstracts (25 abstracts per journal each year) of articles published in the *British Medical Journal* (*BMJ*), the *Canadian Medical Association Journal* and the *Journal of the American Medical Association* (*JAMA*) in 1988 and 1989 (nonstructured abstracts) and in 1991 and 1992 (structured abstracts).
Main outcome measures: The quality of abstracts was measured against 33 objective criteria, which were divided into eight categories (purpose, research design, setting, subjects, intervention, measurement of variables, results and conclusions). The quality score was determined by dividing the number of criteria present by the number applicable; the score varied from 0 to 1.
Results: The overall mean quality scores for nonstructured and structured abstracts were 0.57 and 0.74 respectively ($p < 0.001$). The frequency in meeting the specific criteria was generally higher for the structured abstracts than for the nonstructured ones. The mean quality score was higher for nonstructured abstracts in *JAMA* than for those in *BMJ* (0.60 v. 0.54, $p < 0.05$). The scores for structured abstracts did not differ significantly between the three journals.
Conclusions: The findings support recommendations that promote the use of structured abstracts. Further studies should be performed to assess the effect of time on the quality of abstracts and the extent to which abstracts reflect the content of the articles.

Objectif : Évaluer et comparer la qualité des résumés non structurés et structurés d'articles de recherche originaux publiés dans trois journaux médicaux.
Conception : Étude d'observation à l'insu fondée sur des critères.
Échantillon : Échantillon aléatoire de 300 résumés (25 résumés par journal par année) d'articles publiés dans le *British Medical Journal* (*BMJ*), le *Journal de l'Association médicale canadienne* et le *Journal of the American Medical Association* (*JAMA*) en 1988 et 1989 (résumés non structurés) et en 1991 et 1992 (résumés structurés).
Principales mesures de résultats : On a mesuré la qualité des résumés en fonction de 33 critères objectifs répartis en huit catégories (objectif, conception, contexte, sujets, intervention, mesure des variables, résultats et conclusions). On a établi la note de qualité en divisant le nombre de critères présents par le nombre de critères pertinents; le résultat s'est établi entre 0 et 1.
Résultats : Les résultats globaux moyens de qualité des résumés non structurés et structurés se sont établis à 0,57 et 0,74 respectivement ($p < 0,001$). En général, les auteurs ont observé les critères plus souvent dans les résumés structurés que dans les résumés non structurés. Les résumés non structurée parus dans le *JAMA* ont obtenu une note de qualité moyenne plus élevée que ceux qui ont paru dans le *BMJ* (0,60 c. 0,54, $p < 0,05$). On n'a pas constaté d'écart

From the Faculty of Pharmacy, University of Toronto, Toronto, Ont.

Reprint requests to: Dr. Thomas R. Einarson (attention Ms. Anna Taddio), Faculty of Pharmacy, University of Toronto, 19 Russell St., Toronto, ON M5S 2S2; fax (416)978-8511

← For prescribing information see page 1707

important, entre les trois journaux, quant aux résultats obtenus par les résumés structurés.
Conclusions : Les constatations appuient les recommandations favorables à l'utilisation de résumés structurés. Il y aurait lieu de procéder à d'autres études pour voir s'il se produit avec le temps une évolution de la qualité des résumés et évaluer dans quelle mesure les résumés reflètent le contenu des articles.

Improving the quality of scientific literature has been advocated since the inception of scientific publication, in 1665.[1] In the early 1960s the *Journal of the American Medical Association* (*JAMA*) took an innovative step toward improving biomedical communication by moving the summary and conclusions of articles to the beginning.[2] The *Canadian Medical Association Journal* (*CMAJ*) also adopted this abstract format.[3] Given that there are over 2 million biomedical articles published[4] annually in over 20 000 biomedical journals,[5] extracting pertinent literature has become formidable. Abstracts enable readers to determine more quickly which articles are of interest to them.[6]

As reported by Didolkar, Flemming and Venanzi[7] the American Medical Association stated that "there is no intrinsic difference between a well-prepared summary and an abstract." Abstracts should provide all the necessary and important information on the research performed (e.g., the study's purpose, design, results and conclusions).[8] They enable readers to review relevant features of the research without having to read the entire report. However, investigators have demonstrated that abstracts can be misleading or biased and that the entire article should still be read.[9–11]

Typically abstracts present information in four general sections: the introduction, methods, results and conclusions.[12] In 1987 a proposal by an ad hoc working group for further improvement of the abstract format was published in the *Annals of Internal Medicine*.[8] The group advocated a structured format for abstracts of research articles, with the following sections: objective, design, setting, patients or participants, interventions, measurements and main results, and conclusions. The structured format was proposed to make literature searches and literature evaluations more accurate.[8] Structured abstracts unmask methodologic problems that were left out in traditional abstracts.[12] They also facilitate more consistent peer review.[13,14]

As important as the abstract is to readers, it is surprising how little effort has been put into them.[6] Studies of the quality of scientific abstracts offer editors and readers insight into the shortcomings of abstracts and how they compare with abstracts in other journals. They can also facilitate improvements in current standards of practice.

Recently, investigators have begun to assess the quality of abstracts in biomedical journals. Using objective criteria Narine and associates[15] assessed the quality of traditional abstracts of the 33 original research articles published in *CMAJ* in 1989. The mean overall score of the abstracts was 0.63 out of 1. Abstracts were found to be deficient in the reporting of technical descriptors of study design, study variables and subject selection. Results were often reported without supporting data, and many abstracts failed to address study limitations and recommendations for future study.

Comans and Overbeke,[16] using criteria from the Ad Hoc Working Group,[8] analysed structured abstracts of original research articles published in three major medical journals: the *British Medical Journal* (*BMJ*), the *New England Journal of Medicine* and the *Annals of Internal Medicine*. They found that the abstracts were clearly written but often lacked information about sample selection, patient demographics and statistical analyses. Froom and Froom,[17] in studying whether structured abstracts published in the *Annals of Internal Medicine* conformed to the same published guidelines, found deficiencies in similar areas such as patient selection, number of refusers, number of dropouts and reasons for dropping out. Haynes[18] pointed out that a comparison of nonstructured and structured abstracts in similar journals would have been more informative.

As a follow-up to these studies[15–17] we set out to assess and compare the quality of nonstructured and structured abstracts appearing in *BMJ*, *CMAJ* and *JAMA* over 4 selected years. The specific objectives of this study were (a) to assess the quality of abstracts of original research articles in three medical journals before and after the implementation of the structured format, (b) to compare the overall quality of nonstructured and structured abstracts and (c) to determine whether the quality of nonstructured and structured abstracts differed between the three journals.

Methods

The journals selected for the study were the *BMJ*, *CMAJ* and *JAMA* because they are the principal journals of national medical associations and require an abstract for all original research articles. The 4 years studied were 1988, 1989, 1991 and 1992. The instructions to authors of the three journals were examined over the 4 study years to detect any differences in instructions regarding abstracts. Abstracts published in 1988 and 1989 were mainly nonstructured and thus represented the pool of traditional abstracts available for quality assessment; structured abstracts were excluded. Abstracts published in 1991 and 1992 were mainly structured and thus represented the pool of structured abstracts available for study; nonstructured ones were excluded. *BMJ*,[1] *CMAJ*[19] and *JAMA*[20] officially adopted the structured format in 1988, 1990 and 1991 respectively. Since structured abstracts began to appear in all three journals in 1990, we

viewed that year as being transitional and excluded it from the study.

Only abstracts of original research articles were eligible. They were placed in chronologic order by journal. A random sample of 25 abstracts per year for each journal was selected with the use of a computer-generated random-number list. A total of 300 abstracts were studied. Abstracts were photocopied and numbered using a computer-generated four-digit random-number list. The abstract citations and assignment numbers were transcribed onto paper for future identification. The abstracts were then retyped by two research assistants in a standardized format without the journal citation, author(s), study setting or other markers that would reveal the identity of the journal. The abstract format was not changed. During the assessments the abstracts were identified only by their assigned numbers in an attempt to eliminate bias.

The 32 evaluation criteria were taken from the study by Narine and associates[15] and separated into their eight categories. We added one criterion: Was any information regarding the level of statistical significance or other relevant statistical measures given in the results section? Each criterion was represented by a question on a checklist and could be answered as Yes, No or Not applicable. All responses were weighted equally.

Abstract quality (Q) was defined as the proportion of criteria present according to the formula $Q = Y \div (Y + N)$, where Y was the number of Yes and N the number of No answers. Not applicable responses were omitted. The maximum quality score was 1 and the minimum 0.

Four reviewers evaluated the quality of the abstracts. In a pilot study 10 nonstructured abstracts were reviewed independently, and the consistent interpretation of the criteria was discussed. The second phase of the pilot study involved an independent rating of five nonstructured and five structured abstracts and an assessment of interrater reliability to ensure that the raters concurred in their application of the criteria.

After the pilot study each rater was given 75 randomly selected abstracts to evaluate independently. Five nonstructured and 5 structured abstracts were taken from the 300 abstracts to reassess interrater reliability. Raters recorded quality scores on preprinted grading sheets. If they could not independently rate any aspect of an abstract they would bring it to the group for consensus grading.

Interrater reliability was assessed with a Pearson's product moment correlation coefficient; the method described by Rosenthal[21] was used to correct for attenuation in correlations. The effective interrater reliability was then calculated using the Spearman–Brown formula. The χ^2 test was used to compare the journal formats and the frequency with which specific criteria were met. The Bonferroni correction factor was included to account for multiple comparisons. A p value of 0.05 or less was considered significant.

Results

The instructions to authors varied regarding the maximum number of words allowed in nonstructured abstracts. Otherwise, all three journals instructed authors to follow the "conventional" or "uniform" method of writing abstracts. *BMJ* did not provide detailed instructions on how to write a structured abstract, but *CMAJ* and *JAMA* did.

The pilot test yielded high interrater agreement (r = 0.631 with 8 degrees of freedom [df] [coefficient of variation = 0.039], $p < 0.001$, effective interrater reliability = 0.872). The second test performed during the study produced even higher agreement (r = 0.809 with 8 df [coefficient of variation = 0.058], $p < 0.001$, effective interrater reliability = 0.944).

The types of research articles included for which abstracts were rated did not differ between the three journals (data available on request). The mean quality scores for nonstructured and structured abstracts are shown in Table 1. The overall quality score was higher for structured abstracts than for nonstructured abstracts (0.74 v. 0.57, $p < 0.001$). The scores differed significantly over time: 1991 and 1992 abstracts were each significantly different from the 1988 and 1989 abstracts ($p < 0.05$). The quality scores did not differ significantly for nonstructured abstracts between 1988 and 1989 or for structured abstracts between 1991 and 1992.

JAMA had a significantly higher mean quality score than *BMJ* for nonstructured abstracts ($p < 0.05$). However, no difference was observed in the mean quality scores for structured abstracts between the three journals.

Table 2 shows the frequency with which nonstructured and structured abstracts met each of the 33 criteria measured. The frequency was higher for structured abstracts in 32 (97%) of the criteria; it was significantly higher in 14 (42%). For nonstructured abstracts the frequency was less than 75% for 24 (73%) of the criteria; for structured abstracts this occurred in only 10 (30%) of the criteria.

Table 1: Mean quality scores* for nonstructured and structured abstracts in the *British Medical Journal* (*BMJ*), the *Canadian Medical Association Journal* (*CMAJ*) and the *Journal of the American Medical Association* (*JAMA*) in 1988–89 and 1991–92

	Journal; mean quality score			
Abstract	*BMJ*	*CMAJ*	*JAMA*	Overall score†
Nonstructured‡	0.54	0.57	0.60	0.57
Structured	0.74	0.73	0.74	0.74

*To calculate the score, the number of Yes responses to questions concerning the evaluation criteria were divided by the sum of the Yes and No responses.
†$p < 0.001$.
‡$p < 0.05$, between *JAMA* and *BMJ*.

Discussion

We assessed the quality of a random sample of nonstructured and structured abstracts from three medical journals using a criterion-based survey. The structured abstracts received significantly higher quality scores than the nonstructured abstracts, which suggests that the structured format is preferable to the conventional, nonstructured format in providing complete information.

The higher quality scores for structured abstracts may be a direct result of the design of structured abstracts. The design provides a framework for the information that should be included and prompts the reader in retrieving this information. However, the more comprehensive structured abstracts may reduce the likelihood that a person will read the entire article. This can be a problem, because abstract quality may not reflect the quality of the article; the abstract was not intended to replace reading of the article and should not be used for this purpose. Structured abstracts presumably assist readers in literature searches, but this has not yet been substantiated.

The following information was reported more frequently with the structured format than with the non-

Table 2: Proportion of abstracts that met the study criteria, by type of abstract

Criterion	Type; % of abstracts		No. of applicable abstracts*
	Nonstructured	Structured	
Purpose			
Was any information on the purpose given?	92.0	100.0†	150/150
Was the purpose explicitly stated?	73.6	97.3†	140/150
Was the main purpose distinguished from secondary ones?	48.7	95.2†	37/83
Research design			
Was any information on the research design given?	78.7	98.7†	150/150
Were technical descriptors used?	45.6	88.4†	125/146
If a follow-up study, was the duration given?	71.8	77.4	39/53
Setting			
Was any information on the setting given?	61.7	94.6†	149/148
Was the level of clinical care (e.g., primary care) indicated?	60.5	90.0†	81/100
Subjects			
Was any information on the subjects given?	93.8	99.3	146/143
Were common demographic characteristics given?	14.0	29.0	136/138
Were technical descriptors of subject selection (e.g., random sample) used?	14.9	43.6†	141/140
Was the number of subjects indicated?	82.6	91.6	144/143
Were the response and refusal rates indicated?	11.3	25.5	97/106
Was the number of dropouts and losses indicated?	4.9	26.0†	102/104
If the samples were matched were matching characteristics given?	40.0	88.9	15/9
Intervention			
Was any information on intervention given?	88.5	98.3	122/115
Were the commonest name and common synonyms given?	66.7	81.8	30/33
Was a description given?	57.1	79.5†	112/112
Was the duration indicated?	46.3	44.6	54/56
Measurement of variables			
Was any information on the measures given?	96.7	99.3	150/149
Were the variables explicitly defined?	45.6	80.5†	147/149
Was the source of the data given?	62.4	77.2	149/149
If the measurements were subjective were the observers blinded to the patient groupings?	13.8	19.3	29/31
Results			
Were any results given?	97.3	100.0	150/150
Were they directly related to the purpose?	94.3	99.3	123/150
Were appropriate numeric data given?	61.9	88.6†	147/149
Were appropriate statistical values given?	12.6	46.2†	135/143
Conclusions			
Were any conclusions drawn?	71.1	98.0†	149/150
Were they directly related to the purpose?	68.5	81.1	89/143
Were they consistent with the results?	79.6	83.3	103/144
Were the study's limitations mentioned?	6.7	10.1	149/139
Were the study's implications mentioned?	48.0	58.7	150/150
Were there recommendations for further study?	8.0	15.7	150/140

*The number of nonstructured abstracts (first number) and structured abstracts (second number) to which the criterion was applicable.
†$p < 0.05$.

structured one: study purpose, setting, number of dropouts, interventions, study variables, appropriate numeric and statistical values, and conclusions. Although the structured abstracts more consistently met the assessment criteria, they did not meet all of them. For both abstract styles, there appears to be room for improvement in this regard. Imperfect quality scores for structured abstracts may be the result of a lack of space allotted for abstracts in medical journals (usually limited to 250 words) and the inability of investigators to concisely summarize their research.

The overall quality score for nonstructured abstracts did not differ significantly between 1988 and 1989. The same was true for structured abstracts between 1991 and 1992. This suggests that the increase observed in the quality score for structured abstracts was mainly due to the structure format and not to time. However, we cannot rule out the effects of time completely, because we studied the quality of each abstract style for only 2 consecutive years. Future research can address the effect of time through the inclusion of abstracts for more than 2 consecutive years for each type of abstract.

We studied the quality of abstracts of original research. However, abstracts are used to describe other types of articles as well. Hayward and associates[22] recently proposed guidelines for writing abstracts of articles on clinical practice guidelines. It is conceivable that the quality of these types of abstracts would be amenable to study. Future studies should address the quality of abstracts of other types of articles and include other journals and disciplines.

In conclusion, structured abstracts provide a more informative summary of research than nonstructured abstracts. We recommend that journal editors and peer reviewers routinely assess the quality of abstracts of original research articles and provide authors with comments on how to modify them if they do not meet an acceptable standard. As a service to investigators writing research articles, journals should include information on how to write structured abstracts. Medical journals should establish a minimum acceptable value and publish abstracts that meet specified criteria.

References

1. Lock S: Structured abstracts: now required for all papers reporting clinical trials. [editorial] *BMJ* 1988; 297: 156
2. Soffer A: Abstracts of clinical investigations. *Chest* 1987; 92: 389–390
3. Squires BP: Structured abstracts. *Can Med Assoc J* 1990; 142: 703
4. Bernier CL, Yerkey AN: *Cogent Communication: Overcoming Information Overload*, Greenwood Press, Westport, Conn, 1979: 39
5. Warren KS (ed): Selective aspects of the biomedical literatue. In *Coping With the Biomedical Literature: a Primer for the Scientist and the Clinician*, Praeger Publishers, New York, 1981: 17–30
6. Pitkin RM: The importance of the abstract. [editorial] *Obstet Gynecol* 1987; 70: 267
7. Didolkar MS, Flemming MV, Venanzi WE Jr: Abstract renders the summary superfluous. *Surg Gynecol Obstet* 1989; 168: 259–262
8. Ad Hoc Working Group for Critical Appraisal of the Medical Literature: A proposal for more informative abstracts of clinical articles. *Ann Intern Med* 1987; 106: 598–604
9. Gotzsche PC: Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Controlled Clin Trials* 1989; 10: 31–56
10. Evans M, Pollock AV: Trials on trial. *Arch Surg* 1984; 119: 109–113
11. Pocock SJ, Hughes MD, Lee RJ: Statistical problems in the reporting of clinical trials. *N Engl J Med* 1987; 317: 246–232
12. More informative abstracts. [editorial] *J Gen Intern Med* 1989; 4: 76
13. Haynes RB, Mulrow CD, Huth EJ et al: More informative abstracts revisited. *Ann Intern Med* 1990; 113: 69–76
14. Mulrow CD, Thacker SB, Pugh JA: A proposal for more informative abstracts of review articles. *Ann Intern Med* 1988; 108: 613–615
15. Narine L, Yee DS, Einarson TR et al: Quality of abstracts of original research articles in *CMAJ* in 1989. *Can Med Assoc J* 1991; 144: 449–453
16. Comans ML, Overbeke AJ: The structured summary: a tool for reader and author. *Ned Tijdschr Geneeskd* 1990; 134: 2338–2343
17. Froom P, Froom J: Deficiencies in structured medical abstracts. *J Clin Epidemiol* 1993; 46: 591–594
18. Haynes RB: More informative abstracts: current status and evaluation. *J Clin Epidemiol* 1993; 46: 595–597
19. Squires BP: Structured abstracts of original research and review articles. [editorial] *Can Med Assoc J* 1990; 143: 619–622
20. Rennie D, Glass RM: Structuring abstracts to make them more informative. *JAMA* 1991; 266: 116–117
21. Rosenthal R: *Meta-analytic Procedures for Social Research*, Sage Publications, Beverly Hills, Calif, 1984: 55–62
22. Hayward RSA, Wilson MC, Tunis SR et al: More informative abstracts of articles describing clinical practice guidelines. *Ann Intern Med* 1993; 118: 731–737