

Comparing data accuracy between structured abstracts and full-text journal articles: implications in their use for informing clinical decisions

Paul Fontelo, Alex Gavino, Raymond Francis Sarmiento

10.1136/eb-2013-101272

Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine, Bethesda, Maryland, USA

Correspondence to

Dr Raymond Francis Sarmiento
Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine, 38A/B1N, 8600 Rockville Pike, Bethesda, MD 20894, USA; sarmientorr@mail.nlm.nih.gov

Abstract

Background The abstract is the most frequently read section of a research article. The use of 'Consensus Abstracts', a clinician-oriented web application formatted for mobile devices to search MEDLINE/PubMed, for informing clinical decisions was proposed recently; however, inaccuracies between abstracts and the full-text article have been shown. Efforts have been made to improve quality.

Methods We compared data in 60 recent-structured abstracts and full-text articles from six highly read medical journals.

Results Data inaccuracies were identified and then classified as either clinically significant or not significant. Data inaccuracies were observed in 53.33% of articles ranging from 3.33% to 45% based on the IMRAD format sections. The Results section showed the highest discrepancies (45%) although these were deemed to be mostly not significant clinically except in one. The two most common discrepancies were mismatched numbers or percentages (11.67%) and numerical data or calculations found in structured abstracts but not mentioned in the full text (40%). There was no significant relationship between journals and the presence of discrepancies (Fisher's exact p value = 0.3405). Although we found a high percentage of inaccuracy between structured abstracts and full-text articles, these were not significant clinically.

Conclusions The inaccuracies do not seem to affect the conclusion and interpretation overall. Structured abstracts appear to be informative and may be useful to practitioners as a resource for guiding clinical decisions.

Introduction

The peer-reviewed medical literature has grown exponentially. Through mobile devices, previously unavailable clinical resources have now become freely accessible worldwide. These include abstracts and some full-text articles of published studies. Through PubMed/MEDLINE, abstracts can now be conveniently accessed at the point of care.¹⁻² The universal accessibility of abstracts as a knowledge resource in many parts of the world with mobile Internet access has sparked interest in using journal abstracts as evidence-based sources when full-text articles are unavailable to clinicians especially in geographic areas with limited resources.¹⁻³ Hence, it was recently proposed that a clinician-oriented web application formatted for mobile devices such as 'Consensus Abstracts' could be used to search and review multiple concurring abstracts in MEDLINE/PubMed to inform clinical decisions.¹

The abstract is the most often read section of a research article.⁴ Journal abstracts and related bottom-line summaries are appealing because they are easy to read and give a quick digest of the article.¹⁻⁴⁻⁸ As a short summary of the full-text article,⁸ the abstract incorporates a significant amount of information when written well and is frequently the only part perused by the reader.⁹ It is essential therefore that the abstract accurately reflects the contents of the full-text article.⁹ It should be a concise, clear and informative representation of the results or interpretation of the full text.¹⁰

Searching abstracts is convenient because it contains most of the article's relevant keywords.¹¹ Many clinicians continue to depend on journal abstracts in seeking answers to clinical questions despite the increasing availability of full-text articles from online archives like PubMed Central¹² and other similar repositories. Clinicians and other healthcare practitioners also rely exclusively on abstracts due to lack of time to read the full-text article, poor critical appraisal skills or limited access to full-text articles, especially in resource-constrained settings.^{5-7 13-15} But there are inherent problems to this approach because dependence on abstracts alone assumes that it is complete and accurate. Incorrect reporting of data in the abstract could bias the reader and lead to misinterpretation of research findings.¹⁶

Over the years, several efforts have been undertaken to make the abstract more informative and minimise inaccuracies between the abstract and the full text of an article. Most journals have adopted the use of structured abstracts. A structured abstract's distinctly labelled sections enhance comprehension of the article.¹⁷ Compared with the unstructured format, structured abstracts are of higher quality and are preferred for providing informative summaries of published studies.¹⁸⁻²¹ Advantages of using structured abstracts also include: expediting the peer review process prior to publication, assisting health professionals in finding clinically relevant journal articles and conducting more detailed literature searches.^{22 23} In 1987, a standard abstract format for articles reporting clinical studies was introduced.²⁴ This structured abstract format was comparable with the introduction, methods, results and discussion (IMRAD) format.²⁵ In effect, IMRAD has now become a de facto standard for the full text of scientific journal articles and is also commonly used as a structure for journal article abstracts since it follows the course of scientific discovery.²⁶⁻²⁸ The Consolidated Standards of Reporting Trials (CONSORT) group and the International Committee of Medical Journal Editors (ICMJE) recommend the use of structured abstracts although there may be differences in the precise format from one journal to the next.^{29 30} This has led to a steady increase in the



number of journal articles with structured abstracts in MEDLINE citations. A recent study by the National Library of Medicine reported that the proportion of structured abstracts added yearly to MEDLINE increased from 0.4% in 1991 to 23.0% in 2008 and the number of journals publishing structured abstracts expanded from 78 in 1989–1991 to 3166 in 1992–2006 time periods.³¹

However, even with structured abstracts, Pitkin *et al*^P observed that it was common for medical journals to contain data in abstracts that were inconsistent with or absent from the full-text article. They found a persistence of inaccurate data in the abstracts (18–68%) in six general medical journals. Another study discovered that in one medical specialty journal, at least 25% of manuscripts that were returned after revision, comprised of data in the abstracts that were unverifiable in the full text.³² A study by Boutron *et al*³³ found 'spin' (defined as reporting strategies that emphasise the beneficial effect of the experimental treatment even with statistically non-significant primary outcomes or divert the reader from results with no statistical significance) in the results (37.5%) and conclusions (58.3%) sections of abstracts of randomised controlled trial (RCT) reports published in 2006. A study by Yavchitz *et al*¹⁶ recently identified 'spin' in 47% of press releases and media coverage with 'spin' found in abstract conclusions of RCTs as the only factor associated. This is a significant finding because clinicians often rely their initial assessment of a clinical trial on the abstract,²⁹ and in some geographic areas, the abstract may be all that clinicians will have access to.

In this study, we aimed to determine the accuracy of data contained in structured abstracts and their usefulness for guiding clinical decisions, by attempting to answer the following questions: (1) Do the data in the structured abstract accurately reflect that in the full text? Are the numbers reported in the structured abstract the same as in the full text? In this comparison, we concentrated on the numerical data. (2) If there is a discrepancy in the data between the structured abstract and full text, is it clinically significant? (3) Would the structured abstract alone be an adequate guide for clinical decision-making? For items 2 and 3, we tried to evaluate whether data discrepancies found between a structured abstract and its corresponding full text affected the conclusion and clinical 'bottom line' of the published research.

Methods

Between 11 February and 14 March, 2011, 60 articles listed as the 'most read', 'most cited' or 'most e-mailed' were selected from six clinical journals³⁴: *American Journal of Emergency Medicine*, *British Medical Journal*, *Journal of the American Medical Association*, *Lancet*, *New England Journal of Medicine* and *Obstetrics and Gynaecology*. These journals were chosen because they were highly regarded^{34 35} and 'most read' in the medical community.

Using the IMRAD format as a model, we extracted and stored the data in an Excel spreadsheet. If a category did not exactly match the IMRAD format, we assigned it to a cell that closely matched the category. We independently examined the structured abstract for numerical data

and verified its accuracy from the corresponding section in the body of the article, including tables and figures. Differences in assessments were resolved through consensus. Structured abstracts were considered inaccurate if they contained data that were either inconsistent or missing in the full-text article. We did not consider rounded off values in the abstract as discrepancies if the precise value was found in the full text, including tables and figures and if the rounding was done correctly and consistently. We then decided if a discrepancy was: (1) clinically significant or (2) clinically not significant. We defined clinically significant discrepancy as data that would require a clinical response or action if encountered in a clinical setting; was misleading; or would result in any misinterpretation of the conclusions of the study. We also considered data similar to 'spin', as described in a previous work,¹⁶ as clinically significant discrepancy such as misinterpretation of statistically significant results as not significant intervention effects, either neutral or non-beneficial or misinterpretation of statistically not significant results as showing comparable effectiveness or equivalence of treatment.

The proportions of articles containing deficiencies were compared across journals using the Fisher's exact test. We attempted to determine whether there was a significant relationship between the journal that published the article and whether an article contained discrepancies. We also calculated the Fleiss κ coefficient to assess the degree of agreement among the three evaluators. Statistical analyses were performed using R Statistical Software V.2.15.2.³⁶

Results

The six journals examined were high impact journals. At the time of review, their median (IQR) impact factor was 37.99 (4.04–42.033).

Table 1 shows a comparison of the articles with and without discrepancies for each journal. Data inaccuracy was observed in 53.33% of the articles. Analysis showed that there was no significant relationship between the journal that published the article and whether an article contained discrepancies (Fisher's exact p value=0.3405). The Fleiss κ coefficient was 0.869 ($z=12$, $p=0.001$), indicating high agreement³⁷ among the three evaluators.

Table 2 shows the proportion of articles with discrepancies between the structured abstract and the full text by IMRAD sections. The proportion of sections of the structured abstract with discrepancies ranged from 3.33% (Discussion) to 45.00% (Results). Compared to

Table 1 Number and percentages of articles with and without discrepancies (n=60)

Journal	With discrepancy (% of total articles)	Without discrepancy (% of total articles)
A	5 (8.33)	5 (8.33)
B	2 (3.33)	8 (13.33)
C	6 (10.00)	4 (6.67)
D	6 (10.00)	4 (6.67)
E	7 (11.67)	3 (5.00)
F	6 (10.00)	4 (6.67)
Total	32 (53.33)	28 (46.67)

Table 2 Comparison of data accuracy in structured abstracts vs. full texts (n=60)

Section	Discrepancy (%)	Clinically not significant discrepancy (%)	Clinically significant discrepancy (%)	No discrepancy (%)
Introduction	4 (6.67)	4 (6.67)	0 (0.00)	56 (93.33)
Methods	8 (13.33)	8 (13.33)	0 (0.00)	52 (86.67)
Results	27 (45.00)	26 (43.33)	1 (1.67)	33 (55.00)
Discussion	2 (3.33)	2 (3.33)	0 (0.00)	58 (96.67)

other areas, the Results section showed the most number of discrepancies although these were mostly clinically not significant. In only one article (1.67%) was a clinically significant difference found between the structured abstract and full text. This was observed in the Results section that contained the highest percentage of inaccuracies.

Table 3 shows the list of discrepancies found between the structured abstracts and the full texts. Of the 32 articles with discrepancies, we found seven articles (11.67%) with discrepancies in at least two different IMRAD sections. The two most common discrepancies found were clinically not significant. These were numbers or percentages in the structured abstract not matching with those in the full text (11.67%) and numerical data or calculations found in the abstract but not explicitly mentioned in the full text (40%). The

Table 3 Summary of discrepancies found between structured abstract and full text (n=60)

Type of discrepancy	List of discrepancies found in structured abstract	Number of articles with specific discrepancy (%)
Clinically significant discrepancy	Misinterpretation of significant intervention effects	1 (1.67)
Clinically not significant discrepancy	Incomplete listing of study objectives	1 (1.67)
	Setting of study mentioned in abstract but not explicitly mentioned in full text	1 (1.67)
	Incomplete listing of study limitations	1 (1.67)
	Discrepancy in terminology definitions	2 (3.33)
	Inconsistent comparison of methods used	1 (1.67)
	Incomplete explanation of experimental correlations	1 (1.67)
	Inconsistent rounding off of numbers or percentages	2 (3.33)
	Numerical data or percentages did not match	7 (11.67)
	Numerical data or calculations found in abstract but not explicitly mentioned in full text*	24 (40.00)

*We performed recalculations for these data to determine if those found in the abstract matched with the numbers in the full text.

discrepancy pertaining to mismatched numbers or percentages included: mismatched p values, inconsistent range of values and mismatched numbers or percentages of variables. However, the mismatched values did not adversely affect the final interpretation of results and were still considered as clinically not significant. For the discrepancy related to numerical data or calculations mentioned in the abstract but not directly mentioned in the full text, we performed recalculations and observed that the numbers, percentages or calculations found in the structured abstract were consistent with the data presented in the full text had they been calculated.

Discussion

The proportion of discrepancies we found between the structured abstracts and full texts in the six journals examined was quite large (53.33%), but consistent with previous observations by Taddio *et al*¹⁸ and Wong *et al*³⁸. For us, this was surprising considering that these were highly regarded medical journals with a high median impact factor. The discrepancies found were mostly due to inconsistencies in numerical data between the abstract and those in the full text, but were regarded as clinically not significant. We deemed these as minor discrepancies and could be explained either because of a mismatch in data numbers or percentages or missing statistical calculations in the full text but found in the abstract. Recalculations proved that the numbers presented in the structured abstract were indeed correct and consistent with the data presented in the full text. There were no discrepancies attributed to data found only in the abstract. The discrepancies found were not considered to be misleading nor would result in any misinterpretation of the findings on the part of the reader. Overall, we observed that the structured abstracts were still appropriate surrogates¹⁰ of full-text articles.

To practice evidence-based medicine (EBM), clinicians need to critically appraise full-text articles to guide their clinical decision-making.¹ Applying EBM in clinical practice encourages the use of timely and relevant information to complement one's expertise.³⁹ Clinicians look for significant relevant findings in treatment protocols, diagnostic examinations and outcomes of certain interventions. In this study, most of the discrepancies found between the structured abstract and the full text were not significant clinically. The sole article that showed a clinically significant discrepancy was a misinterpretation of a significant intervention effect, explicitly stated as 'HbA1c was significantly lower (p=0.008)' in the Results section of the full-text article but was shown as 'no difference in HbA1c among agents on the intensive group' in the abstract. However, this discrepancy did not have any effect in the article's conclusion and overall interpretation. There were

no clinically significant discrepancies observed in other sections of the article.

Using medical literature to guide clinical decisions has been shown to be effective.^{40–44} To be most useful, current clinical evidence needs to be conveniently accessible.⁴⁵ When provided access to either abstracts or full-text articles, improvements in clinical decisions have been observed.³ With a high percentage of clinicians (69%) using abstracts only in seeking answers to clinical questions,³ it is crucial that a high standard of quality for published abstracts is present. A 10-year follow-up study found that using structured abstracts, regardless of precise formats, has helped maintain a high quality of available abstracts.³⁸

Our study had several limitations. The journal articles selected may not be representative of all scientific articles in the medical literature. The small sample of articles examined limits the generalisability of the findings in this study. We also only evaluated the utility of structured abstracts in guiding clinical decisions. We did not include research articles with unstructured abstracts. More reviewers would enhance the reliability of this type of evaluation. Further analysis with a broader sample of medical journals to include articles with unstructured abstracts is needed.

In this study, we still found a large proportion of mostly clinically not significant inaccuracies between the abstracts and full texts consistent with previous reviews. These discrepancies did not seem to affect the clinical bottom line. This study may provide some support that evidence derived from structured abstracts may be another reliable resource for clinicians. Nevertheless, the high proportion of inaccuracies suggests that there is still room for improvement. Greater attention needs to be devoted by authors, reviewers and editors in improving the quality of abstracts. Clinicians must also remember that one abstract alone may not be enough in making clinical decisions^{46 47}; hence, the need to review many, perhaps using tools that can facilitate such review such as the web application 'Consensus Abstracts'.¹

Limitations The journal articles selected may not be representative of all scientific articles in the medical literature. The small sample of articles examined limits the generalisability of the findings in this study. We also only evaluated the utility of structured abstracts in guiding clinical decisions. We did not include research articles with unstructured abstracts. More reviewers would enhance the reliability of this type of evaluation. Further analysis with a broader sample of medical journals to include articles with unstructured abstracts is needed.

Contributors PF designed the study, participated in analysis of the results and helped draft the manuscript. AG participated in the study design and analysis of the results. RFS conducted statistical and qualitative analysis of the results and drafted the manuscript.

Funding This research was supported by the Intramural Research Programme of the National Institutes of Health (NIH), National Library of Medicine (NLM) and Lister Hill National Center for Biomedical Communications (LHNCBC).

This research was also supported in part by an appointment to the NLM Research Participation Programme, administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the US Department of Energy (DoE) and the NLM.

Disclaimer The views and opinions of the authors expressed herein do not necessarily state or reflect those of the National Library of Medicine, National Institutes of Health or the US Department of Health and Human Services.

Competing interests None.

References

- Fontelo P. Consensus abstracts for evidence-based medicine. *Evid Based Med* 2011;16:36–8.
- Hoogendam A, Stalenhoef AF, Robbé PF, *et al*. Analysis of queries sent to PubMed at the point of care: observation of search behaviour in a medical teaching hospital. *BMC Med Inform Decis Mak* 2008;8:42.
- Marcelo A, Gavino A, Isip-Tan II, *et al*. A comparison of the accuracy of clinical decisions based on full-text articles and on journal abstracts alone: a study among residents in a tertiary care hospital. *Evid Based Med* 2013;18:48–53.
- Winker MA. The need for concrete improvement in abstract quality. *JAMA* 1999;281:1129–30.
- Read MEDLINE abstracts with a pinch of salt. *Lancet* 2006;368:1394.
- Barry HC, Ebell MH, Shaughnessy AF, *et al*. Family physicians' use of medical abstracts to guide decision making: style or substance? *J Am Board Fam Pract* 2001;14:437–42.
- Haynes RB, McKibbon KA, Walker CJ, *et al*. Online access to MEDLINE in clinical settings: a study of use and usefulness. *Ann Intern Med* 1990;112:78–84.
- Doyle LB. Semantic road maps for literature searchers. *J ACM* 1961;8:553–78.
- Pitkin RM, Branagan MA, Burnmeister LF. Accuracy of data in abstracts of published research articles. *JAMA* 1999;281:1110–11.
- Ries JE, Su K, Peterson G, *et al*. Comparing frequency of content-bearing words in abstracts and texts in articles from four medical journals: an exploratory study. *Med Info* 2001;5:381–4.
- Shah PK, Perez-Iratxeta C, Bork P, *et al*. Information extraction from full text scientific articles: where are the keywords? *BMC Bioinformatics* 2003;4:20.
- US National Library of Medicine. PubMed Central. <http://www.ncbi.nlm.nih.gov/pmc> (accessed Jan 2013).
- Saint S, Christakis DA, Saha S, *et al*. Journal reading habits of internists. *J Gen Intern Med* 2000;15:881–4.
- The PLoS Medicine Editors. The impact of open access upon public health. *PLoS Med* 2006;3:e252.
- Berwanger O, Ribeiro RA, Finkelsztein A, *et al*. The quality of reporting of trial abstracts is suboptimal: survey of major general medical journals. *J Clin Epidemiol* 2009;62:387–92.
- Yavchitz A, Boutron I, Bafeta A, *et al*. Misrepresentation of randomized controlled trials in press releases and news coverage: a cohort study. *PLoS Med* 2012;9:e1001308.
- US National Library of Medicine. Structured abstracts. http://www.nlm.nih.gov/bsd/policy/structured_abstracts.html (accessed 15 Nov 2012).
- Taddio A, Pain T, Fassos FF, *et al*. 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. *CMAJ* 1994;150:1611–15.
- Dupuy A, Khosrotehrani K, Lebbé C, *et al*. Quality of abstracts in 3 clinical dermatology journals. *Arch Dermatol* 2003;139:589–93.

20. Sharma S, Harrison JE. Structured abstracts: do they improve the quality of information in abstracts? *Am J Orthod Dentofacial Orthop* 2006;130:523–30.
21. Trakas K, Addis A, Kruk D, *et al.* Quality assessment of pharmacoeconomic abstracts of original research articles in selected journals. *Ann Pharmacother* 1997;31:423–8.
22. Haynes RB, Mulrow CD, Huth EJ, *et al.* More informative abstracts revisited. *Ann Intern Med* 1990;113:69–76.
23. Hayward RS, Wilson MC, Tunis SR, *et al.* More informative abstracts of articles describing clinical practice guidelines. *Ann Intern Med* 1993;118:731–7.
24. 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.
25. Huth EJ. Structured abstracts for papers reporting clinical trials. *Ann Intern Med* 1987;106:626–7.
26. Sollaci LB, Pereira MG. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. *J Med Libr Assoc* 2004;92:364–7.
27. Nakayama T, Hirai N, Yamazaki S, *et al.* Adoption of structured abstracts by general medical journals and format for a structured abstract. *J Med Libr Assoc* 2005;93:237–42.
28. Kulkarni H. Structured abstracts: still more. *Ann Intern Med* 1996;124:695–6.
29. Hopewell S, Clarke M, Moher D, *et al.*; CONSORT Group. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 2008;5:e20.
30. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals: writing and editing for biomedical publication. <http://www.icmje.org> (accessed 14 Dec 2012).
31. Ripple AM, Mork JG, Knecht LS, *et al.* A retrospective cohort study of structured abstracts in MEDLINE, 1992–2006. *J Med Libr Assoc* 2011;99:160–3.
32. Pitkin RM, Branagan MA. Can the accuracy of abstracts be improved by providing specific instructions? A randomized controlled trial. *JAMA* 1998;280:267–9.
33. Boutron I, Dutton S, Ravaud P, *et al.* Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 2010;303:2058–64.
34. McKibbin KA, Wilczynski NL, Haynes RB. What do evidence-based secondary journals tell us about the publication of clinically important articles in primary healthcare journals? *BMC Med* 2004;2:33.
35. Thomson Reuters. Thomson Reuters Announces New 2011 Journal Citation Reports. http://thomsonreuters.com/content/press_room/science/688332 (accessed 10 Oct 2012).
36. R Development Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2012. <http://www.R-project.org> (accessed 5 Jan 2013).
37. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
38. Wong HL, Truong D, Mahamed A, *et al.* Quality of structured abstracts of original research articles in the British Medical Journal, the Canadian Medical Association Journal and the Journal of the American Medical Association: a 10-year follow-up study. *Curr Med Res Opin* 2005;21:467–73.
39. Sackett DL, Rosenberg WM, Gray JA, *et al.* Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312:71–2.
40. McGowan J, Hogg W, Campbell C, *et al.* Just-in-time information improved decision-making in primary care: a randomized controlled trial. *PLoS ONE* 2008;3:e3785.
41. Crowley SD, Owens TA, Schardt CM, *et al.* A Web-based compendium of clinical questions and medical evidence to educate internal medicine residents. *Acad Med* 2003; 78:270–4.
42. Westbrook JI, Coiera EW, Gosling AS. Do online information retrieval systems help experienced clinicians answer clinical questions? *J Am Med Inform Assoc* 2005;12:315–21.
43. Leon SA, Fontelo P. MedlinePlus en Español and Spanish-speakers. *AMLA Annu Symp Proc* 2007;1028.
44. Hersh WR, Crabtree MK, Hickam DH, *et al.* Factors associated with successful answering of clinical questions using an information retrieval system. *Bull Med Libr Assoc* 2000;88:323–31.
45. Sackett DL, Straus SE. Finding and applying evidence during clinical rounds: the 'evidence cart'. *JAMA* 1998;280:1336–8.
46. Tonelli MR, Curtis JR, Guntupalli KK, *et al.* An official multi-society statement: the role of clinical research results in the practice of critical care medicine. *Am J Respir Crit Care Med* 2012;185:1117–24.
47. Kavanagh BP. Glucose in the ICU—evidence, guidelines, and outcomes. *N Engl J Med* 2012;367:1259–60.