Structured Abstracts in MEDLINE: Twenty-Five Years Later

2012

Anna M. Ripple[1], James G. Mork[1], John M. Rozier[2], Lou S. Knecht[3]

[1] ripple@nlm.nih.gov
mork@nlm.nih.gov
Lister Hill National Center for Biomedical Communications
National Library of Medicine
National Institutes of Health

[2] john.rozier@nih.gov
Office of Computer and Communications Systems
National Library of Medicine
National Institutes of Health

[3] knechtl@nlm.nih.gov
Division of Library Operations
National Library of Medicine
National Institutes of Health

BACKGROUND: Structured abstracts, which are abstracts with distinct, labeled sections, are a medical publishing innovation that began in 1987[1]. Twenty-five years later, nearly a quarter of all abstracts added yearly to MEDLINE®/PubMed® are structured. In 2010, structured abstracts were reformatted for display in PubMed for easier readability with bolded, upper case section labels each beginning on a new line.

OBJECTIVE: To characterize the distribution of structured abstracts in journals indexed for MEDLINE.

METHODS: The U.S. National Library of Medicine (NLM) developed and validated a new algorithm for identifying structured abstracts in MEDLINE in 2010[2]. Recently, NLM updated its research on structured abstracts by running the algorithm against the 2012 MEDLINE Baseline (annual, static versions of the MEDLINE data are freely available for use by any researcher from the MEDLINE/PubMed Baseline Repository)[3].

RESULTS: For the 2012 MEDLINE Baseline release, 4,525 journals contributed 1,817,573 structured abstracts. Figure 1 shows a quartile distribution of the concentration of structured abstracts in journal titles indexed in MEDLINE. Seventy-seven journals contributed 25% of the structured abstracts; 258 journals contributed 50%; and 674 journals contributed 75%. The remaining 25% of the structured abstracts is spread across 3,516 journal titles. By the end of 2010, more than 1,600 journals or about 30% of the 5,484 journals indexed in that year routinely publish structured abstracts (starting in or prior to 2008 and continuing through 2010, the last

complete publication year in the 2012 MEDLINE Baseline, while also contributing a total of at least one hundred structured abstracts).

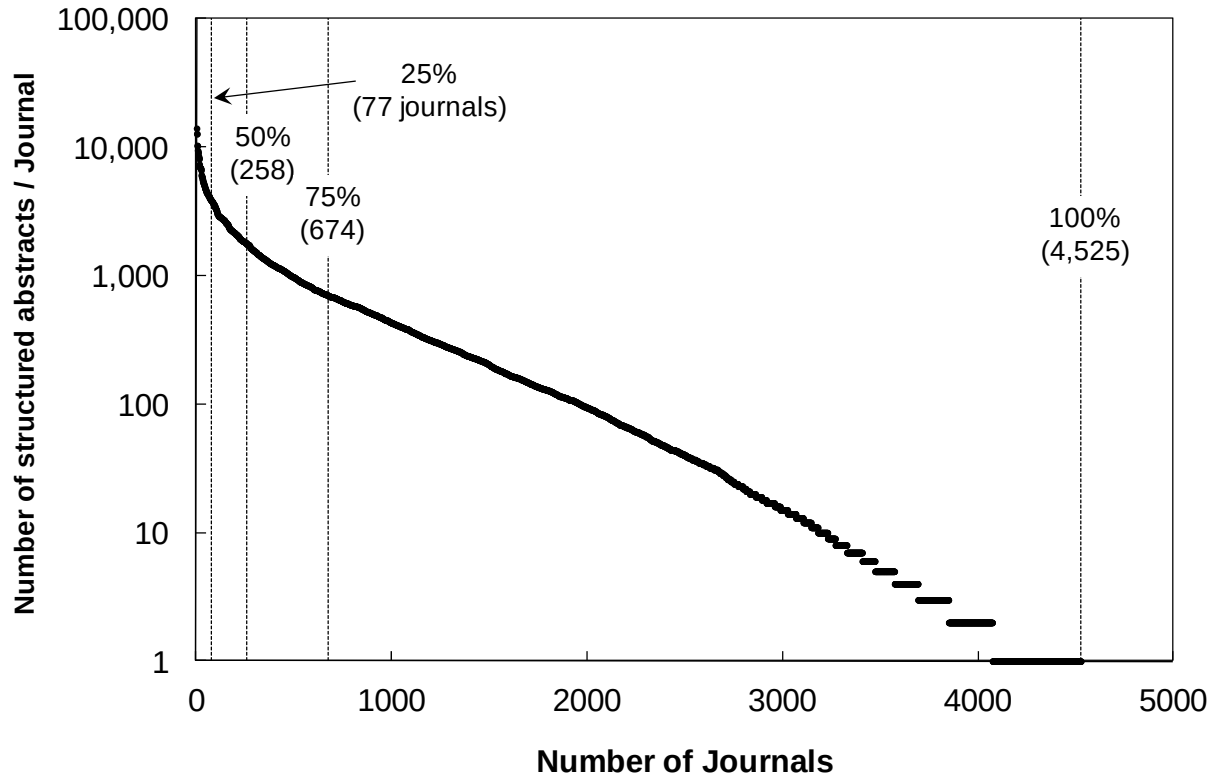

Figure 1: Quartile Distribution of Structured Abstracts across 5,484 Indexed MEDLINE Journals (2010)

The 2012 data show that the average number (mean and mode) of labels in a structured abstract is four. In addition, 69.27% of the citations (1,259,197/1,817,573) contain four labels. The labels tend to be less specific than those in the original structured abstract proposed standard formats[4], e.g., only METHODS rather than DESIGN, INTERVENTION, and PATIENTS. This is consistent with findings from two smaller studies of structured abstracts in clinical journals which revealed that most structured abstracts had either the introduction, methods, results, and discussion (IMRAD) format or a four-label variant that included an initial label introducing the topic of the article (such as BACKGROUND, PURPOSE), middle labels of METHODS and RESULTS, and a final label of CONCLUSION(S)[5,6].

From the 7,791,344 labels identified in the 1,817,573 structured abstracts, NLM isolated about 2,000 unique label phrases, all of which NLM mapped to five general categories: BACKGROUND, OBJECTIVE, METHODS, RESULTS, and CONCLUSIONS.

CONCLUSIONS: The substantial growth in both the individual number of MEDLINE records with structured abstracts and in the number of journals that continuously publish structured

abstracts demonstrates widespread adoption of structured abstracts over the past twenty-five years. NLM has implemented the label category maps as part of the annual NLM data dissemination. NLM is continuing to identify and map labels on an annual basis. These mappings are freely available from the NLM Web Resource called Structured Abstracts in MEDLINE[3]. NLM is also exploring the utility of structured abstracts in assisting the indexing process, and enhancing information retrieval and discovery. Datamining, in particular, may benefit from targeting certain labels for specific reasons, e.g., RESULTS may yield better information for gene/disease relationships than BACKGROUND.

REFERENCES

1. Huth EJ, Case K. Annals of Internal Medicine at age 75: reflections on the past 25 years. Ann Intern Med. 2002 Jul 2;137(1):34-45.

2. Ripple AM, Mork JG, Knecht LS, Humphreys BL. A retrospective cohort study of structured abstracts in MEDLINE, 1992-2006. J Med Libr Assoc. 2011 Apr;99(2):160-3.

3. US National Library of Medicine. Structured abstracts in MEDLINE® [Internet] Bethesda, MD: The Library; [rev. 1 May 2012; cited 27 July 2012]. < http://structuredabstracts.nlm.nih.gov>

4. Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. More informative abstracts revisited. Ann Intern Med. 1990 Jul 1;113(1):69-76. Review.

5. Nakayama T, Hirai N, Yamazaki S, Naito M. Adoption of structured abstracts by general medical journals and format for a structured abstract. J Med Libr Assoc. 2005 Apr;93(2):237–42.

6. Kulkarni H. Structured abstracts: still more. Ann Intern Med. 1996 Apr 1;124(7):695–6.

.

ACKNOWLEDGMENTS

Conflict of Interest Disclosures: All authors have no Conflicts of Interest to report.
Funding Support: This research was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC).