Ubiquitous but Unfinished:
Online Information Retrieval Systems

William R. Hersh, MD

When I first began research on information retrieval (IR) systems in the late 1980s,[1] I could not have imagined their growth in the ensuing decade and a half. At that time, the few physicians who used them were called "early adopters"[2] and virtually no one fathomed the notion of searching by patients or consumers. Now, in the early 21st century, IR systems are ubiquitous, and everyone involved in health care, from patients to physicians to policy makers, is using them.[3] A good deal of credit goes to the US National Library of Medicine, which has made an entry way into the biomedical literature free and easy through its MEDLINE PubMed system (even if many of the articles themselves are not free) and also made searching easier for laypeople through its MEDLINEplus system.

Despite their ubiquity, there are still research challenges for IR systems, especially in the area of evaluation. There are still unanswered questions about how to deploy IR systems at the point of care and evaluate their efficacy. Past research does give us some guideposts. We know, for example, that physicians have frequent and unmet information needs.[4,5] We know that searching the primary literature takes way too long—upwards of a half an hour per question to find, read, and critically appraise articles.[6] In addition, we know that relying on journal abstracts, though much improved in recent years, may obscure essential details in full papers.[7] We also know that textbooks and other synopses may lag behind the cutting edge of science.[8]

Nonetheless, the busy clinical environment makes it imperative that clinicians have access to highly synoptic, aggregated information. Ideally, that information should be derived from the best evidence. This still begs the question of how useful such systems are in clinical care and whether clinicians can use their information accurately. Several studies over the years have shown that although clinicians can access information about a topic with relative ease, they do not always find the best information, nor do they answer their questions reliably.[6,9]

There are also many other challenges in evaluating IR systems. One can lay out the spectrum of evaluation as a continuum from the feasibility level (e.g., do the computer algorithms work?) to the controlled laboratory setting (e.g., can users in a simulated environment use the system successfully?) to the real world (e.g., does the system improve outcomes of care?). As one goes to the higher levels, evaluation becomes more challenging, not only because real users must be inconvenienced in the studies but also because it is difficult to isolate the impact that the IR system has on the outcome of an episode of care. Although it is fairly straightforward to measure the impact of a decision support system that recognizes prescribing errors, determining the value of consulting an IR system in a patient's ongoing congestive heart failure is more challenging. For this reason, IR system evaluations have tended to focus either on the system, using a test collection of "canned" queries and a fixed document collection, or on the user in a simulated laboratory setting.

The study in this issue by Westbrook and others is an important advance in the evaluation of IR systems.[10] Although it is a controlled laboratory type of study, it uses a large and diverse mix of Australian physicians who have access to a wide variety of clinical information resources in an integrated system. Probably the most important result of this study verifies what was shown in an earlier study of medical and nurse practitioner students using MEDLINE,[6] which was that despite the ease of use and ubiquity of these systems, only 50% of the questions are answered correctly. Although the IR system is better than no system at all, considerable improvement is still required.

The study also had some interesting findings concerning the confidence users had in their answers. In general, those who had the correct answers were more confident, indicating that those who knew the correct answer were more able to discern that they obtained it. However, a significant fraction of users persisted in feeling confident about an answer that was incorrect.

Address correspondence to William R. Hersh, Department of Medical Informatics & Clinical Epidemiology, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Mailcode BICC, Portland, OR 97239; e-mail: hersh@ohsu.edu.

These results show that even if the IR systems we build work perfectly, getting users to identify the correct answers to questions from their use is still a challenge.

Although this study provides a good assessment of clinicians using IR systems in a simulated setting, we do not know how generalizable these results are to real-world searching. Are the questions developed for the experiment indicative of those encountered in real clinical practice? Even if the topics are indicative, do clinicians ask them in the way they are asked here? Even further, what is the significance to the quality of clinical care if clinicians can only answer a question, even using a state-of-the-art information tool, 50% of the time? Does this impact the quality of care? Does it lead to medical errors?

If nothing else, these questions demonstrate the importance of continued research on IR systems. Baseline research has demonstrated beyond a doubt that the knowledge base in the head of the average physician is woefully incomplete. But we cannot automatically assume that information resources, whether in print or electronic form, will automatically improve things. Continued research must assess the information needs of clinicians, the best evidence to meet those needs, and how that information is most effectively delivered to improve the quality of care.

REFERENCES

1. Hersh WR, Greenes RA. Information retrieval in medicine: state of the art. MD Comput. 1990;7:302–11.

2. Marshall JG. Characteristics of early adopters of end-user online searching in the health professions. Bull Med Library Assoc. 1989;77:48–55.

3. Fox S, Fallows D. Internet health resources: health searches and e-mail have become more commonplace, but there is room for improvement in searches and overall Internet access. Washington (DC): Pew Internet & American Life Project; 2003. Available from: http://www.pewinternet.org/pdfs/PIP_Health_Report_July_ 2003.pdf

4. Gorman PN, Helfand M. Information seeking in primary care: how physicians choose which clinical questions to pursue and which to leave unanswered. Med Decis Making. 1995;15:113–9.

5. Ely JW, Osheroff JA, Ebell MH, et al. Analysis of questions asked by family doctors regarding patient care. Br Med J. 1999;319:358–61.

6. Hersh WR, Crabtree MK, Hickam DH, et al. Factors associated with success for searching MEDLINE and applying evidence to answer clinical questions. J Am Med Informat Assoc. 2002;9:283–93.

7. Haynes RB, Mulrow CD, Huth EJ, et al. More informative abstracts revisited. Ann Intern Med. 1990;113:69–76.

8. Antman EM, Lau J, Kupelnick B, et al. A comparison of results of meta-analyses of randomized controlled trials and recommendations of clinical experts: treatments for myocardial infarction. JAMA. 1992;268:240–8.

9. Hersh WR, Hickam DH. How well do physicians use electronic information retrieval systems? A framework for investigation and review of the literature. JAMA. 1998;280:1347–52.

10. Westbrook JI, Gosling AS, Coiera EW. The impact of an online evidence system on confidence in decision making in a controlled setting. Med Decis Making. 2005;25(2):178–185.