Collaborative Information Synthesis

Catherine Blake
Department of Information and Computer Science, University of California, Irvine, CA 92697-3425
E-mail: cblake@ics.uci.edu

Wanda Pratt
Information School and Division of Biomedical & Health Informatics, University of Washington, Seattle, WA 98195-2840. E-mail: wpratt@u.washington.edu

As the quantity of scientific literature continues to soar, scientists struggle to keep up with new findings, even in narrow areas of expertise. Although advances in information retrieval have eased the task of finding relevant articles, scientists now must face the challenge of aggregating information from within the retrieved set of documents. Our study explores the user behavior and information requirements of scientists as they interact with medical literature to answer research questions. We found that although their information needs were clearly defined, they still refined the retrieval, extraction, and analysis phases of a process that we have called information synthesis. We also found that they actively collaborated throughout the process. We describe their behavior and introduce our design and progress towards our tool METIS (Multi-user ExTraction and Information Synthesis) that will support the collaborative information synthesis process used by public health and biomedical scientists.

1 Introduction

The recent increase in electronically available articles holds great promise for scientists in medicine and public health, who have a long tradition of using the results of previous studies to gain insight into new research questions. However access to scientific text is a necessary, but not sufficient criterion to answer research questions to advance science. It is rarely the case that a single article will answer a research question completely. Instead, scientists must integrate information from multiple studies, which differ in sample size, populations, interventions, and often in their conclusions. We consider information access within a larger framework that we call information synthesis. **Information synthesis** is a process used to answer a research question by integrating findings from multiple documents. Using an iterative process, scientists retrieve documents, extract the required information from those documents and analyze the extracted details.

Our research explores (1) user behaviors exhibited by scientists in medicine and public health as they use biomedical literature to achieve their research goals, (2) the information requirements implied by their work practices, and (3) the development and evaluation of information technology that supports the user behavior and information requirements identified in 1 and 2. Our goal is to enable scientists to incorporate new findings reported in medical literature, faster and more comprehensively than the current manual approaches allow. To gain the insight necessary to achieve this goal we conducted a user study to understand the context surrounding the information synthesis process.

To illustrate the inadequacies of the manual approach currently employed to integrate findings, we provide an overview of the **systematic review** process, an instance of the information synthesis process. A recent survey indicated that the time between an initial idea for a systematic review and its later publication is 28 months (Petrosino 1999)[1]. To make the scientist's job even more difficult, during a typical 28-month period, the National Library of Medicine adds approximately 933,000 articles to MEDLINE, the premier citation database for public health and medicine (NLM 2002). Although not all these articles will relate to one review, they contribute to the scientist's information overload problem. For example, if we conducted a systematic review to evaluate the effect of breast cancer risk attributable to smoking, it would be difficult but paramount to consider the approximately 12,600 articles that are typically published on breast cancer during the 28 months required to conduct the review. To reduce this information overload problem, scientists often reduce the number of articles included in their analysis by constraining the research question. However, this reduction could introduce undesirable biases, and thus reduce the validity of the result. Such delays and biases could directly impact public health. As the number of articles published each year continues to increase, the current manual

---

[1] Based on 14 lead authors who conducted reviews as part of the Cochrane Collaboration. Participants are generally not paid for their time, which might cause this estimate to be slightly exaggerated.

approaches will become even more difficult to use than they already are.

Our study is a first step towards both understanding the information synthesis process and developing tools to assist scientists with that process. We present both our characterizations of user behavior during information synthesis, and our design and implementation progress towards a **Multi-user ExTraction and Information Synthesis (METIS)** system that will support the user behavior we observed.

2 Related Work

Many researchers have investigated information needs in the domain of medicine. Studies have reported the information behavior of a variety of people, such as physicians (Smith 1996), nurses (Lange 1993), and even patients (Leydon, Boulton et al. 2000) in settings as diverse as rural clinics (Bowden, Kromer et al. 1994) and academic medical centers (Woolf and Benson 1989). Researchers have used various methods, such as observations (Forsythe 1998), surveys (Williamson, German et al. 1989), and interviews (usually based on the critical-incident technique) (Northup, Moore-West et al. 1983). Although we have learned much about information behavior in these clinical settings, little work has been done to understand the information needs of public health and medical scientists in a research setting. Our work examines this important user group for the task of information synthesis.

Researchers in information science have studied the general process of information seeking. Although our findings are consistent with much of that research, we focus more on the use of the information than most models have. Our findings are most consistent with those of Leckie & Pettigrew (Leckie and Pettigrew 1997), Ellis (Ellis 1989) (particularly his characteristic of extracting as part of the search process), and Kulthau (Kuhlthau 1991). However, we focus much less on the affective aspects of Kulthau's six-stage seeking process, because emotions are less apparent in work settings, and we delve deeper into the last two steps from her model: the collection and presentation of the information, in this case, for the particular task of information synthesis.

Collaboration with respect to integrating information is an aspect of information behavior that the research community has largely neglected. Most previous research has focused on collaboration from the perspective of the person with an information need collaborating with an experienced searcher (such as a librarian) to address the first individual's need (Fowell and Levy 1995). In contrast, we observed collaboration among equally experienced members of medical research teams. Systems have been developed to support this type of collaboration (Procter, Goldenberg et al. 1998), but few support teams that collaborate to search for as well as use documents. A few exceptions are for systems that support collaborative browsing among many individuals (Nichols, Pemberton et al. 2000; Twidale and Nichols 1998). Their focus is on supporting the browsing process by allowing collaborators to see a trace of all the documents that users' visited. In the setting that we study, the research question, and thus, their information need, is well defined, and therefore more appropriate for searching, rather than such browsing activities. Recently, the National Science Foundation funded a joint project with members from the University of Washington, Microsoft, Boeing, and Risø National Laboratory to explore collaborative information seeking, but their focus is on engineering and software-development teams (Fidel, Bruce et al. 2000).

From a computer-science perspective, the information synthesis process is a type of multiple document summarization. Several computer-science researchers have developed multi-document summarization systems (Mani and Bloedorn 1999; McKeown and Radev 1995). These systems use natural language processing techniques to extract representative sentences and connect them into a textual summary, or to generate new sentences for the text summary based on the content of representative sentences. However, scientists in public health and medicine care little about the natural-language sentences and paragraphs; they are more interested in extracting specific, quantitative, and qualitative values from the text and the techniques used to integrate that extracted information.

We are aware of one other approach that proposes to support information synthesis in a medical setting (Sim 1998; Sim, Owens et al. 2001). In their model, authors would submit the results of a clinical trial as both a database entry to a "trial bank" in addition to the published article. This approach has not yet considered how to incorporate material published before the trial bank or how such an approach would work with different types of studies, such as cohort studies.

Although researchers have explored the development of extraction techniques, few consider the context or behavior of users who would use this technology. Exceptions include such as the Alembic Workbench (Day, Aberdeen et al. 1997) and a system that interactively created rules to extract information from text (Caruana, Hodor et al. 2000). The Automatic Content Extraction (ACE) program supports the development of extraction techniques that operate on news articles, speech and images sources (Doddington 2000). Motivated by the human genome project, scientists have developed techniques to extract biological information such as, protein names (Fukua, Tsunoda et al. 1998), protein structures (Humphreys, Demetriou et al. 2000), and molecular pathways (Friedman, Kra et al. 2001). In the clinical domain, the Medical Language Extraction and Encoding System (MedLEE) was developed for radiology reports (Friedman, Alderson et al. 1994) and later applied to identify suspicious mammography findings (Jain and Friedman 1997).

3 Data Collection Methodology

We base our characterizations of user behavior on three sources of information: documents, formal training, and observations and interviews of scientists in public health and medicine.

3.1 Document Analysis

During our observations and interviews, we actively read material about the systematic review process. Our primary sources of information were documents produced by the Cochrane Collaboration (Clarke and Oxman 2001) and **Health Technology Assessment (HTA)** (Sutton, Abrams et al. 1998). The **Cochrane Collaboration (CC)** is a world leader at conducting and providing access to systematic reviews related to biomedical literature (http://www.cochrane.org). We supplemented these documents with books (Hunter and Schmidt 1990; Ingelfinger, Mosteller et al. 1994; Lipsey and Wilson 2000; Petitti 2000) and articles (Bartolucci 1999; Davies and Crombie 1998; DerSimonian and Laird 1986; Engels, Schmid et al. 2000; Petrosino 1999; Stroup, Berlin et al. 2000).

Recommendations of the information that should be included in a study that has synthesized information provided us with insight into the types of information that scientists require. We found recommendations from three working groups that described the type of information that should be included in three kinds of articles used. The Asilomar Working Group provided a checklist for reports of clinical trials, (Asilomar Working Group, 1996). The second group, QUOROM, recommended information that should be included in a medical meta-analysis (Moher, Cook et al. 2000; Moher, Cook et al. 1999) and public health meta-analysis (Stroup, Berlin et al. 2000). The last group, CONSORT made recommendations for information reported for randomized clinical trials (Begg, Cho et al. 1996),(Moher, Schultz et al. 2001).

3.2 Formal Training

To accelerate our familiarity with the systematic review process, the first author attended a presentation and short course. The presentation given at the University of California, conducted by Dr. Adams, was an open lecture that discussed systematic reviews. The short course focused on meta-analytic techniques and was conducted during a half-day by Dr. Olkin as part of a Decision Making Conference in San Diego (Olkin 2001).

3.3 Observations

The main component of our study of researchers' information behavior centered on our observations of them at work. We identified two groups of scientists who use medical literature extensively in their work. The first group we refer to as the **medical group**, who performed their study at the Susan Samueli Center for **Complementary and Alternative Medicine (CAM)**. The second group we refer to as the **public health group**, who worked for the Health Priorities Research Group in the School of Social Ecology at the University of California, Irvine. We now describe the tasks undertaken by each group, and the techniques we used to observe their information behavior.

3.3.1 Medical Group

We began observing the medical scientists at their first organizational meeting in June 2001. The groups' purpose was to conduct non-biased, rigorous reviews of medical literature related to complementary and alternative approaches to medicine. We did not attend, but have access to the minutes of the search strategy meetings. The first author attended the *methodology* meetings, where the group defined the information required from each article. E-mail was used to distribute minutes of each meeting.

With the exception of the bio-statistician, the group had established membership before we began observing their behavior. Group membership was comprised of two experts in the area of spinal manipulation (the focus of the review), a medical librarian, and four experts in other areas of complementary and alternative medicine. This team satisfied their pre-defined criteria to include expertise in the areas of methodology, information science, biostatistics, health services, clinical research, and clinical content. One of the spinal manipulation experts had extensive experience with the systematic review process.

The group conducted a systematic review to answer the research question *What is the reliability of spinal palpatory procedure(s)?* Although it was clear that the review would include spinal manipulation, coming up with the specific research question within the area of spinal manipulation was a collaborative exercise. Other candidate questions included *What is the effectiveness of osteopathic spinal manipulation for low back pain?* and *What is the validity of spinal palpatory procedure for screening and diagnosis of patients with spinal neuro-muscular dysfunction?* The group has planned to conduct a review on the latter research question after they complete the pending reliability review, which demonstrates the iterative nature of the information synthesis process.

In addition to defining the research question clearly, the group also decided the kind of analysis that they would perform. During the very early stages, the group considered using a **meta-analysis,** a quantitative, statistical approach to synthesize information from multiple studies. However, their analysis was refined to a qualitative systematic review because there were an insufficient number of studies that met the criteria for a meta-analysis. This demonstrates the iterative nature of the information synthesis process.

The experts in spinal manipulation provided the initial search terms to initiate the retrieval process. The librarian-contributed expertise was to identify multiple databases and manipulate the search string to use on each. Eleven databases were used to collect citations: Pubmed MEDLINE, Ovid Mantis, CINAHL, EmBase, Web of Science, OCLC PaperFirst, Biosis Preview, Index to

Chiropractic Literature, PEDro (Physiotherapy Evidence Database), Cochrane Library and MD consult. For each database, the librarian e-mailed citation details, including the title and abstract to the entire group for review. During weekly meetings, group members provided additional search terms to the librarian and exchanged articles. Once the group was satisfied that they had identified all relevant citations, they augmented the citations by manually searching the references and by requesting additional recommendations from experts in the field. The group provided each external expert with a complete bibliography of citations collected thus far, and asked them to recommend additional literature, including **gray literature**, work that has not yet been published. A combination of electronic and manual search strategies were necessary to ensure the identification of all potentially relevant articles.

Towards the end of the search process, the medical group began face-to-face discussions and interactions via e-mail to define what information they would extract from each article. They used samples from other spinal manipulation reviews, the CC, HTA, and their own expertise in the spinal manipulation and alternative medicine to identify the information required for each article. To verify that the extraction requirements were clearly stated, three group members and a spinal-manipulation expert who had not been involved with the development of the extraction rules, used the rules to extract information from three articles. They were satisfied that the extraction rules were clear because they all extracted similar information. The people would extract information for each study. Once each person had extracted the details, they would compare results and if the extracted item differed, each group member would explain their rationale. If the differences could not be resolved within the sub-group, a fourth person would consider the evidence and make the final decision. This verification phase indicates the importance of high precision during extraction.

In addition to a qualitative analysis, the group assigned a quality score to each article. This score captured the presence or absence of information and a weight indicating the importance of each information element. For example, the information *Examiners were blinded to clinical presentations* had a quality score of eight, while reporting the age or ethnicity of patients had quality scores of one. The quality score was the sum of each information element.

3.3.2 Public Health Group

Scientists in public health conduct research that correlates peoples' behavior or public policy to diseases, as opposed to medical group, which typically concentrate on finding better treatments or diagnostic tests. This public health group already had experience at extracting information to obtain data for their database of life-saving metrics and cost effectiveness (Tengs, Adams et al. 1995). The first author observed their behavior by working in the same office for two days a week during the summer of 2001. Although we observed behavior for other projects that required the extraction of information from articles, we based our analysis on a completed meta-analysis that explored the relationship between smoking and impotence.

The first author discussed the impotence study with the director, statistician, and research assistants involved in the retrieval of articles and extraction of information. This group designed a search strategy to reduce the effect of publication bias. Specifically, they considered all studies on impotence, and then a research assistant manually identified articles that reported tobacco use, a process that took approximately two person months (N. Williams, Personal Communication). A good systematic review will report the article selection criteria. This study considered 1008 articles that (1) had either "impotence" or "erectile dysfunction" in the title; (2) were not reviews, letters, comments, editorials, or news; (3) involved human, not animal, subjects; (4) were published in 1980 or later; and (5) that clinical trial that were performed in the United States (Tengs and Osgood 2001).

After the articles were retrieved a different group member read each article and recorded (1) the total number of impotent subjects, (2) the number of impotent men who were current smokers, (3) the definition of impotence used in the study, (4) the definition of smoking used in the study, (5) mean, standard deviation, and range of ages in the subject population, (6) the geographic location of the study, (7) the time period over which assessment occurred, and (8) whether the article mentioned smoking in the abstract. (ibid.). This manual process took approximately three hours per study (T.Tengs, Personal Communication).

Another group member, who was not involved in the retrieval or extraction phases preprocessed data from the **Behavioral Risk Factors Surveillance System (BRFSS)** (BRFSS 2001). This data was used to identify the smoking rate of men with a similar age distribution, year and geographic location of each article.

To put the current manual techniques in context consider exploring the relationship between smoking and breast cancer using a search approach adopted by the public health group. A search for documents in MEDLINE for articles that (1) have a keyword of breast neoplasms; (2) are not a review article or letter; (3) involve human subjects; and (4) were published since 1980, yields approximately 70,000 articles. If it takes the same amount of time to identify breast cancer articles that reported tobacco usage, as impotence articles, then it would take 11.5 person years to identify the breast cancer articles that are required for the study. If a similar proportion of breast cancer articles report tobacco use as impotence articles, it would take approximately 1.8 person years to extract the information from these studies. Based on the average annual number of articles produced over the last ten years, approximately 72,000 articles on breast cancer would be published during the 13.3 person years it would take to perform our review. It is not surprising that no one has conducted this review, despite the potential impact of the result.

3.4 *Interviews*

During the preliminary meetings, members from both groups answered many questions regarding the IS process in general, the development of a search string, the data-extraction and analysis methodologies. In addition to participating at meetings and seeking clarification, the first author conducted a 20-minute interview with the expert in spinal manipulation from the medical group (who had previous meta-analysis experience), and a 45-minute interview with both the primary author and the statistician from the public health group. We sent additional questions to the statistician via e-mail and discussed via telephone the details of the BRFSS preprocessing. To determine an estimate of the amount of time they took to search, we discussed the search process with the staff member who manually searched the articles.

4 The Information Synthesis Process

Our study resulted in a detailed picture of the information-synthesis process in the context of scientists in medicine and public health. We learned that these scientists need and use explicit, rigorous, non-biased, and repeatable methods to identify, extract, and analyze information. Although these traits are the cornerstone of a systematic review of biomedical literature, they clearly are true of good scientists in other disciplines.

As shown in figure 1, the research question is central to the information synthesis process. Although the strategy for answering the research question was refined during the process, we observed that the research question drove the requirements of the retrieval, extraction, and analysis phases. The research question asked by the public health and medical groups were *What is the link between smoking use and impotence?* and *What is the reliability of spinal palpatory procedures?*

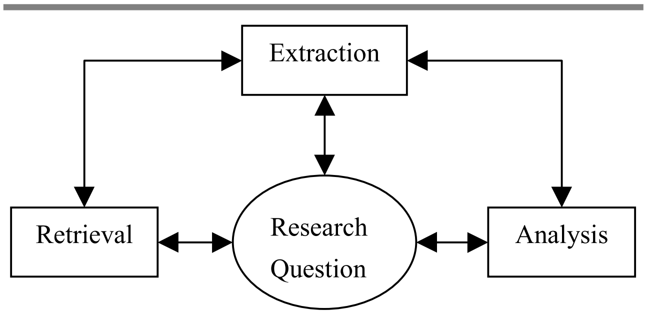

Figure 1 – The Information Synthesis Process. Scientists collect articles, extract information from those articles, and analyze the extracted information. The result of each phase enables scientists to refine the search criteria and extraction requirements. The availability of information influences the type of analysis conducted.

In contrast to our information synthesis model, Davis and Crombie have defined a 5-step process that describes the systematic review process as (1) defining the appropriate question (2) searching the literature (3) assessing the retrieved studies (4) combining study results, and (5) placing the findings in context (Davies and Crombie 1998).

During the retrieval phase, scientists require all potentially relevant articles; thus, recall is critical. Scientists use multiple databases and search strategies to collect possibly relevant citations, however we observed that they also augment the electronically available citations with hand-searched citations (those taken from the references of an article) and gray literature (i.e., articles that have been written but not yet published). Gray literature is obtained by asking experts in the domain. This process emphasizes the importance of recall during information synthesis and informal collaboration with experts.

Although scientists should consider all articles that relate to the research question, this does not imply that all articles are included in the analysis. In addition to the search strategy, scientists develop inclusion criteria and exclusion criteria. For example, the public health group considered all articles that contained “impotence” or “erectile dysfunction" in the title, however articles were only included in the study if they reported the smoking rate. Both groups included the search and inclusion criterions in their final manuscript.

Towards the end of the retrieval phase, the medical group developed a methodology to measure article quality. This score, which is based on the presence of information and a weight associated which that information, is used to categorize articles during the analysis. For example, scientists may find differences between findings depending on their low, medium or high quality.

After scientists develop the data extraction form, they use a manual process to ensure that the information extracted is accurate. Both groups verified the information extracted before use. The context of a document was important for this step. For example, many tables in scientific studies adjust for factors such as age and gender. When extracting information from such a table, group members consider these adjustments to ensure accuracy. This process ensures that the required precision is obtained during the extraction phase of IS.

We observed that scientists refined the extraction criterion after performing preliminary analyses. For example, the second group constrained articles to only those from the United States, after it became clear that only those studies would be included in the final analysis. This decision had consequences both to the generality of the research question and to the studies that would be included. From an information retrieval perspective, this corresponds to a change in retrieval criterion.

The last step in the process is to analyze the information and present a coherent summary of the articles that were included in the review. The medical group, who are still preparing their final report, have currently adopted a qualitative approach. They will compare and contrast the quality and outcome of articles. The public health group adopted a quantitative approach and conducted a random effects meta-analysis (DerSimonian and Laird 1986), after pre-processing both the information extracted from articles and the data retrieved from the BRFSS. The choice of analysis technique is influenced by other phases of the IS process. Specifically, the medical group adjusted their analysis due to a limited number of documents, the public health group made adjustments based on the information within the documents.

5 Functional Requirements

We have characterized the user behavior that we observed by the level of collaboration and by the iterative refinements that the scientists made during each phase of the information synthesis process. In addition to access, a system must support these behaviors if it is to enable scientists to incorporate new findings reported in medical literature, faster and more comprehensively than the current manual approaches allow.

In addition to the behavior exhibited by group members, we found regularities in the type of information used by each group of scientists. We also identified regularities in the location of this required information within an article. These regularities suggest that automating the extraction phase of the IS process is possible. We provide a categorization of the facts required by scientists during IS in section 5.3.

5.1 Collaboration

The first user behavior that we observed was that group members actively collaborated throughout the information synthesis process. This behavior was particularly apparent in the medical group, where each group member contributed distinct but complementary skills.

We were curious to see if other systematic reviews of medical literature involved a team approach. To answer this question, we collected all meta-analyses published in MEDLINE between 1990 and 1996 in the top medical journals[2]. Of the 147 studies that satisfied this criterion, 132 had multiple authors. This large portion of multi-authored papers supports our finding that a systematic review of biomedical literature is a collaborative exercise.

In addition to recognizing that users collaborate, it is equally important to identify how and on which tasks this collaboration takes place. We refer again to figure 1 to describe the collaborative activity during each phase of IS.

[2] Based on a previous empirical analysis of meta-analysis (Engels, Schmid, et al., 2000). we used the Annals of Internal Medicine, Archives of Internal Medicine, British Medical Journal, Circulation, Journal of the American Medical Association, Lancet and New England Journal of Medicine.

(1) Developing the Research Question

Although both groups had a general area of focus, the process of defining the specific research question to be answered was highly collaborative. Users held face-to-face meetings to discuss and refine candidate research questions. Group members from the medical group often bought published articles and books to clarify issues discussed during the previous meeting. The clinical experts played a major role during the formulation of the research question to ensure that it did not duplicate a previous study and addressed an area that was important from a clinical perspective.

(2) Retrieval

In the medical group, the librarian and experts in spinal manipulation collaborated to ensure that the retrieval phase was comprehensive. They updated search strings and database choices and incorporated feedback from all group members. Recall that three group members read each article. The first group member to read the article hand searched the references to provide the medical librarian with additional citations. The librarian then obtained and distributed the new article to the group members. In this way, the expertise of the domain expert was coupled with the retrieval expertise of the librarian to ensure that all relevant articles were collected.

In the public health group, the primary investigator and a research assistant collaborated to develop a search string that would ensure that they retrieved all relevant articles. Unlike the area of complementary and alternative medicine, MEDLINE contains much of literature in public health. Another group member collected the full text of each citation from MEDLINE from the University libraries.

We also observed a passive collaboration between the group members and researchers from previous studies. Specifically authors of a systematic review make their search strategy explicit in their published manuscript. In this way, group members collaborated with authors of previously published systematic reviews to develop their search strategy. Recommendations from the HTA and CC were also considered.

(3) Extraction

A recommendation of what should be included in a systematic review provided us with insight into the kind of information to extract. We found that scientists use either their own previous studies, similar studies from the literature, or external sources (such as the HTA or CC) to create a document that specifies what data should be extracted, the **extraction form**. These extraction forms include a combination of general and specific information types. For example, scientists who are conducting a systematic review of the relationship between tobacco use and breast cancer must extract general information, such as the subjects' age and location of the study, in addition to review-specific information, such as their smoking rate.

The medical group developed the data extraction form and the weights assigned to each unit of information collaboratively. Experts in spinal manipulation and the statistician contributed most these decisions, which they discussed at length.

A critical part of assessing the quality of each article was verifying that the extracted information. Although group members often extracted the information independently, the accuracy was verified collaboratively. Members from the medical group resolved differences using discussion. If a difference could not be resolved, a fourth group member would hear the justifications for each response and make the final decision. The public health group conducted a similar verification process.

(4) Analysis

The methodology used to combine results depends on the type of analysis conducted. A systematic review is either qualitative or quantitative. In the case of a quantitative review, a meta-analysis is often used. A meta-analysis weights the contribution of each article based on the variance in the study (The details of this analysis are beyond the scope of this paper). Scientists conducting a qualitative evaluation will typically present the results with respect to the quality of an article. Although the bio-statistician played an important role defining the statistical integration of information, determining the control items was a collaborative process. For example, there are many known risk factors associated with impotence. The public health group selected a control population using age, geographical location and the date of the study. Each of the information items was used when combining the studies to obtain insight into the over-all effect of tobacco usage and impotence.

Both groups collaborated during the development of the manuscript to place the findings in context. In addition to partitioning articles into included and excluded categories based on the inclusion criteria; the medical group also collected background articles. These articles discussed similar research questions, but did not satisfy the inclusion criteria. The researchers used those articles to put the findings of the study in context. Collaboration took part by writing and editing the final manuscript. Experts in other areas of complementary and alternative medicine ensured that a general reader (as opposed to an expert in spinal manipulation) could also understand the final manuscript.

5.2 Retrieval, Extraction and Analysis Refinement

Although the users we studied had well-defined information needs, we observed a continuous refinement of the retrieval, extraction and analysis phases of the IS process. Scientists documented this evolution, particularly the search strategy, because tradition requires that they provide these details when communicating the results of the synthesis to other scientists.

One artifact that reflected this refinement during retrieval was the database of citations. In the medical group, the medical librarian supplied the title and abstract of each article to each domain expert who would read the abstract and propose additional query terms. She would then collect additional articles obtained using these terms and repeat the process. Despite a well-defined search strategy, the development of the queries used to retrieve citations still required refinement.

The scientists also refined the type of analysis performed based on the availability of information. As the public health group explored the relationship between smoking and impotence, they noticed that many articles also reported alcohol consumption. Alcohol consumption was not added to the final analysis because an insufficient number of articles reported the combination of smoking and alcohol, i.e. alcohol and tobacco were considered independently in the source articles. Similarly, the medical group had considered analyzing the articles using a meta-analysis, however there were insufficient randomized clinical trials related to their research question to conduct this analysis, so they analyzed the studies using a systematic review. Studying how users refine the analysis based on the availability of information is only possible when both of these phases are considered as part of a larger process. The IS model offers this context.

We also observed an interesting interaction between the retrieval and extraction phases in the public health group. The public health group required all studies on impotence where the article also reported current tobacco usage. Had they been able to specify this selection criterion as part of their query then, they would not have had to identify those articles manually. This suggests that the retrieval and extraction phrases of the IS process can be tightly coupled.

5.3 Information Requirements

In addition to how scientists use the medical literature, we were surprised to find that regularities in the type of information required by public health and biomedical scientists during information synthesis. Certain information extracted for the analysis of impotence and smoking was also required for the analysis on spinal manipulation. Further, that data extraction forms available from the HTA and CC had overlapping information requirements. We identified four categories of information required for IS, which we base primarily on the public health study (see Figure 2). Although the specific details extracted will of course vary based on the actual review, we anticipate that these characteristics will generalize to other reviews of medical literature. We have categorized the types of information required by scientists for information synthesis into four categories: information related to the study, population group, intervention or risk factor and the medical condition. We include this list to demonstrate that regularities in the type of information required for information synthesis exist. We ordered the categories in increasing specificity to a review.

The location of information within an article has implications to indexing, retrieval and extraction strategies. We found that information used during information synthesis is often only located within the body of a document, rather than in the title or abstract that are often used to index articles. We also observed that several facts required for information synthesis are located within the first or second tables within a document, such as the age and gender of subjects. This suggests that extraction algorithms could exploit the regularities within a document.

1. Information related to the study

- Number of subjects with medical condition (e.g., number of patients who are impotent)
- Year of publication and that data was collection
- Geographical location of the study (e.g., city, state, country)

2. Information related to the population group

- Gender of participants (e.g., female, male)
- Ethnicity
- Age of participants (including mean, average, standard deviation where available)

3. Information related to the intervention or risk factor

- Details of the intervention or risk factor (e.g., kind of palpation or type of tobacco (cigarettes, pipe))
- Amount of exposure to the intervention or risk factor (e.g., number of palpation's, time smoked)
- Confounding factors related to intervention or risk-factor (e.g., alcohol consumption)

4. Information related to the medical condition

- Location of condition (e.g., cervical, thoracic, lumbar)
- Severity of disease (e.g., mild, moderate or severe pain)
- Confounding factors related to other medical conditions (e.g. heart disease, ovarian cancer)

Figure 2 – Information Requirements of Information Synthesis. Despite the variation in the types of articles used by different groups, we found regularity in the type of information that scientists extracted from those articles. Items 1 and 2 appear to generalize to other areas of medicine, while items 3 and 4 are specific to a particular study.

6 METIS Design

We call our system METIS because in Greek mythology, the titaness Metis presided over all wisdom and knowledge, and we hope METIS will help scientists preside over the wisdom and knowledge in their field. We now describe our vision of the METIS system and our progress towards achieving full functionality to support the behavior that we observed during the information-synthesis process (Figure 3 provides a graphical representation of functionality.)

6.1 Design Overview

METIS is an integrated retrieval and extraction system to provide support for the information behaviors that we defined in Section 5. Our goal is to implement a system that will enable scientists to perform information synthesis of biomedical literature more effectively and more comprehensively than current manual approaches allow. Our long term goal is for scientists to use this semi-automated approach to conduct reviews on research questions that current manual processing techniques have only partially answered because manual techniques were too time consuming.

The system is comprised of three distinct, but integrated components: retrieval, extraction, and analysis. Based on the user behavior that we observed, METIS also supports user collaboration and iteration between the different components, which we describe further in Sections 6.2-6.4.

Because a central finding of our study was that group members actively collaborated throughout the information synthesis process, we decided to include basic support for such collaboration in the system. The collaboration that we observed was synchronous and group members were co-located; however, we do not believe that these constraints are required to perform information synthesis of biomedical literature.

The METIS design centers on providing remote, asynchronous access to artifacts used throughout the information synthesis process. These **artifacts** include the citation database, full-text articles (both related to the methodology and the actual study), data extraction forms, a database of extracted information, meeting minutes, and drafts of the final manuscript.

Simply providing access to artifacts does not support the refinement between retrieval, extraction, and analysis phases. We are constructing an **interactive extraction** component that will enable multiple users to verify information that the system has automatically extracted. This component will provide users with the context of the document. Although some aspects of the process appear well suited to remote, asynchronous collaboration, we believe that this system would also support face-to-face meetings. We will evaluate the interactive component of the system by working with scientists to explore the relationship between smoking and breast cancer.

6.2 Retrieval

As you might expect from the name Multi-user ExTraction and Information Synthesis, our work focuses more on extraction and analysis, and less on retrieval. Retrieval is reflected in the design because the recall performance during this phase is critical to the validity of the IS process. To support the iterative process that we observed, METIS must enable users to refine their retrieval strategy after they have performed later phases of the information synthesis process.

We have modularized the tasks of collecting citations and the full text of articles. Although we anticipate that the current trend toward electronic publishing will continue, it is inevitable that articles required for analysis (such as those in the gray literature) will not be available electronically.

We partitioned the citation details from the queries to reduce redundancy. Scientists have access to multiple citation databases (such as those used by the medical group related that we described in section 3.3.1) and there is inevitably overlap between these sources. We plan to explore CORBA technology to access citation databases in future versions. The combination of the citation database and access to full text would enable automation of searching references, and removal of duplicates that scientists currently perform manually. Scientists could also use this database to add citations in their final manuscript.

Reporting the query used is an important component of the systematic review process, so scientists already make sure that they accurately record the query string used. METIS will follow this habit, and the trend in information retrieval systems that enable queries to be saved, such as Wiley Interscience interface. We anticipate that scientists will be required to search multiple databases. Thus, future versions of METIS will enable scientists to store and refine queries to multiple databases, and maintain citation information from those databases.

The METIS component that retrieves citations from MEDLINE for a given search criterion is complete. This module, written in Java uses the Application Programmers Interface (API) provided by the National Library of Medicine. Augmenting this code with the query constraints

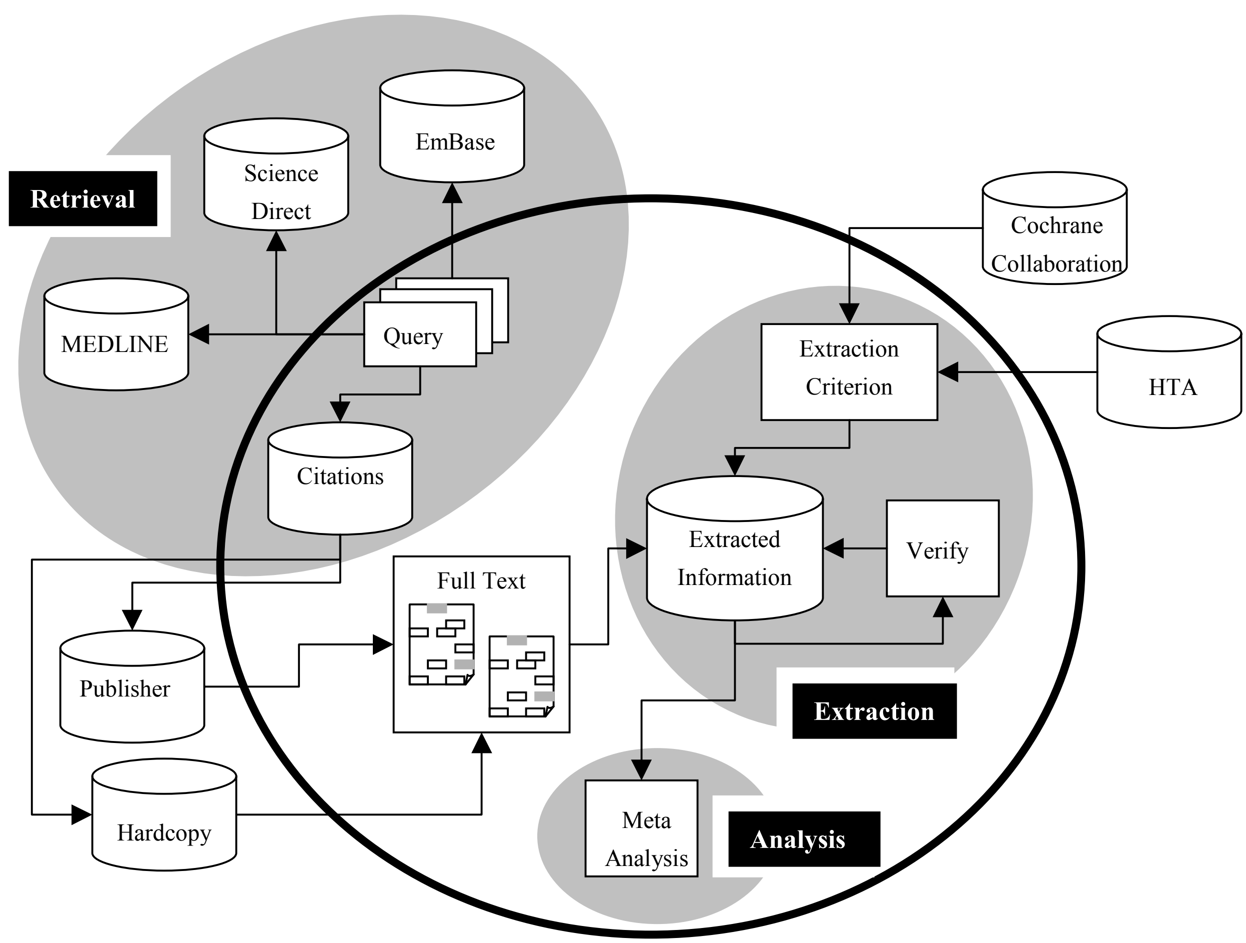

Figure 3 – METIS System Architecture. METIS is an integrated retrieval and extraction system that enables scientists to perform a systematic review of biomedical literature more effectively than the current manual approaches. METIS extracts information from within the full text of articles, and then enables users to verify that information. The system then enables the information to be analyzed using a random effects meta-analysis.

available from the CC to retrieve only those articles from MEDLINE that are suitable for a medical meta-analysis should be simple (Clarke and Oxman 2001). However, these constraints do not apply to our public-health meta-analysis on smoking and breast cancer; thus, we have not included it in our current implementation.

Based on our observation that information required by scientists is located within a document, it is critical that METIS has the full text of biomedical articles available, rather than only the title and abstract. The current pseudo-standard format for peer-reviewed articles is PDF. We have implemented a METIS component that translates PDF files into HTML and from HTML into ASCII text. Our motivation for building this module was that we have PDF and HTML files available, and we wanted to transform both into a standard representation for the extraction algorithms.

Our plan is that the METIS system will pre-process articles as they are downloaded and store the results locally. The system will show the local copy to users during the extraction process, because the document context is required to verify the extracted information.

6.3 Extraction

A major focus of our work has been to develop techniques that automatically extract information required for information synthesis directly from text. We focused on this task because we observed that scientists extract several kinds of information that are independent of the study they are conducting, such as information related to the study design and population groups. Although other kinds of information, such as the treatment or risk factor and the disease are specific to the study, we are exploring the use of a medical language system to make these extraction algorithms more general.

In addition to understanding the types of information required for information synthesis, we also gained insights into regularities in the way that authors described this information in an biomedical article. This has enabled us to develop algorithms based on heuristics to extract information from text. We have completed the METIS extraction components that extract the age of patients, their smoking status, gender, location of the study, and year that the study took place. We are in the process of evaluating these algorithms on 884 breast cancer articles that we have available in electronic form.

To support the collaboration that we observed on the verification task, METIS will enable multiple users to check that the information extracted is accurate. A completely automated extraction process would be unsuccessful because accuracy is critical in public health and medicine domain. We will build on techniques and interfaces developed by the information extraction communities, We are currently reviewing the interface of this component with experts in breast cancer.

6.4 Analysis

The METIS system will enable scientists to verify the information extracted, as required by the current manual processes. The system would provide users with access to the document content during the verification phase. Scientists must also determine the appropriate transformations, such as converting an odds-ratio to a relative-risk, because these transformations are specific to the research question.

Basic summary information, such as a list of the information extracted from each will enable scientists to refine extraction and retrieval requirements. This information will enable scientists to determine controls during the analysis. For example, the public health group would have detected the lack of alcohol information reported.

Lastly, the METIS system will provide a visual representation of the final analysis. We have completed a METIS component to conduct a random-effects meta-analysis (DerSimonian and Laird 1986). We chose this type of meta-analysis because it accounts for heterogeneity and empirical analysis suggests that it is more conservative than other meta-analysis approaches (Ellison, Zhang et al. 2001).

7 Conclusion

The goal of our study was twofold: to gain an understanding of how scientists in public health and medicine use the biomedical literature to answer research questions, and to design a system that supports those user behaviors and information requirements. We found that the process used by scientists incorporates retrieval, extraction and analysis phases, which researchers have typically studied in isolation. The information synthesis model that we have proposed provides a model to explore interactions between these phases.

We found that that users actively collaborated throughout the information synthesis process, and our design of the METIS system reflects this finding. Other findings that heavily influenced our system design were that scientists need both automated and manual mechanisms to perform a comprehensive search of documents related to the research question, and to verify the accuracy of the information extracted.

We observed that information retrieval is just one phase of a larger process, that we call information synthesis, and that scientists iterate between the retrieval, extraction, and analysis phases throughout this process. We also found that although scientists have clear, well-defined information need, they were unable to avoid the problem of information overload. Further that scientists refined their retrieval strategies. To support the user behavior that we observed, and thus, to reduce information overload, we propose that a system also needs to support the extraction and analysis phases of the information synthesis process.

This paper reports a first step towards understanding how scientists use literature in the in the context of their work and the implications toward a system design that would support their user behavior and information requirements. We will continue to work with scientists as we complete the implementation and evaluation of METIS, a system for Multi-User Extraction and Information Synthesis.

8 Acknowledgements

We are deeply grateful to the Susan Samueli Center for CAM for allowing us to observe their systematic review process and to the HPRG for enthusiastically answering our many questions. We would like to thank Madhu Reddy for helpful pointers to literature. A Multiple Investigator Grant from the University of California, Irvine helped fund this project.

9 References

Alsilomar Working Group (1996) Checklist of information for inclusion in reports of clinical trials. The Asilomar Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. *Ann Intern Med;124(8),*741-3

Bartolucci, A. (1999). The Significance of Clinical Trials and the Role of Meta-Analyses. *Journal of Surgical Oncology* 72: 121-123.

Begg, C., M. Cho, S. Eastwood, R. Horton, D. Moher, et al. (1996). Improving the Quality of Reporting of Randomized Controlled Trials: The CONSORT Statement. *JAMA* 276: 637-9.

Bowden, V. M., M. E. Kromer and R. C. Tobia (1994). Assessment of physicians' information needs in five Texas counties. *Bull Med Libr Assoc* 82(2): 189-96.

BRFSS (2001). Behavioral Risk Factor Surveillance System (BRFSS), Survey data, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services. http://www.cdc.gov/nccdphp/brfss/.

Caruana, R., P. G. Hodor and J. Rosenberg (2000). High Precision Information Extraction. KDD-2000 Workshop on Text Mining,

Clarke, M. and A. D. Oxman (2001). Cochrane Reviewers' Handbook 4.1.1. The Cochrane Library. Oxford.

Davies, H. T. O. and I. K. Crombie (1998). What is a systematic review? *Hayward Medical Communications*.

Day, D., J. Aberdeen, L. Hirschman, R. Kozierok, P. Robinson, et al. (1997). Mixed-Initiative Development of Language Processing Systems. Fifth Conference on Applied Natural Language Processing, Washington,

DerSimonian, R. and N. Laird (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* 7: 177-88.

Doddington, G. R. (2000). CE Automatic Content Extraction. Presentation at the TIDES program Kick-off meeting. Available at:

Ellis, D. (1989). A behavioural approach to information retrieval design. *Journal of Documentation* 45(3): 171-212.

Ellison, R. C., Y. Zhang, C. E. McLennan and K. J. Rothman (2001). Exploring the Relation of Alcohol Consumption to Risk of Breast Cancer. *American Journal of Epidemiology* 154(8): 740-747.

Engels, E. A., C. H. Schmid, N. Terrin, I. Olkin and J. Lau (2000). Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Statistics In Medicine* 19(13): 1707-1728.

Fidel, R., H. Bruce, A. M. Pejtersen, S. Dumais, J. Grudin, et al. (2000). Collaborative Information Retrieval (CIR). Information Seeking in Context 2000: The Third International Conference on Information Seeking and Use in Different Contexts, Goteborg, Sweden, Graham Taylor,

Forsythe, D. (1998). Using ethnography to investigate life scientists' information needs. *Bulletin of Medical Library Association* 86(3): 402-409.

Fowell, S. and P. Levy (1995). Developing a New Professional Practice: A Model for Networked Learner Support in Higher Education. *Journal of Documentation* 51(3): 271-280.

Friedman, C., P. Alderson, J. Austin, J. J. Cimino and S. B. Johnson (1994). A general natural-language text processor for clinical radiology. *Journal of American Medical Informatics Association* 1(2): 161-74.

Friedman, C., P. Kra, M. Krauthammer, H. Yu and A. Rzhetsky (2001). GENIES: a natural-langauge processing system for the extraction of molecular pathways from journal articles. *Bioinformatics (suppl1)*: S74-82.

Fukua, K., T. Tsunoda, A. Tamura and T. Takagi (1998). Toward Information Extraction: Identifying protein names from biological papers. Pacific Symposium on Biocomputing, 707-718.

Humphreys, K., G. Demetriou and R. Gaizauskas (2000). Two Applications of Information Extraction to Biological Science Journal Articles: Enzyme Interactions and Protein Structures. Proceedings of the Pacific Symposium on Biocomputing (PSB-2000), Hawaii, 505-516.

Hunter, J. E. and F. L. Schmidt (1990). Methods of meta-analysis: correcting error and bias in research findings. Newbury Park, CA, Sage.

Ingelfinger, J. A., F. Mosteller, L. A. Thibodeau and J. H. Ware (1994). Biostatistics in clinical medicine, McGraw-Hill Inc.

Jain, N. L. and C. Friedman (1997). Identification of findings suspicious for breast cancer based on natural language processing of mammogram reports. AMIA, Philadelphia, 829-833.

Kuhlthau, C. (1991). Inside the search process: Information seeking from the

users perspective. *Journal of the American Society for Information Science* 42(5): 361-71.

Lange, L. L. (1993). Information Seeking by Nurses During Beginning-of-Shift Activities. Proceedings of the AMIA Fall Symposium, p. 317-321.

Leckie, G. J. and K. E. Pettigrew (1997). A general model of the information seeking of professionals: Role theory through the back door? Information Seeking in Context: Proceedings of an International Conference on Research in Information Needs, Seeking and Use in Different Contexts., Tampere, Finland., 99-110.

Leydon, G. M., M. Boulton, Clare Moynihan, A. Jones, J. Mossman, et al. (2000). Cancer patients' information needs and information seeking

behaviour: in depth interview study. *British Medical Journal (BMJ)* 320: 909-913.

Lipsey, M. W. and D. B. Wilson (2000). Practical Meta-Analysis, Sage.

Mani, I. and E. Bloedorn (1999). Summarizing Similarities and Differences Among Related Documents. Advances in Automatic Text Summarization. I. Mani and M. T. Maybury, The MIT Press**:** 357-379.

McKeown, K. and D. R. Radev (1995). Generating Summaries of Multiple News Articles. Proceedings, 18th Annual International Conference on Research and Development in Information Retrieval, 74-82.

Moher, D., D. J. Cook, S. Eastwood, I. Olkin, D. Rennie, et al. (2000). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUORAM statement. *British Journal of Surgery* 87: 1448-54.

Moher, D., D. J. Cook, S. Eastwood, I. Olkin, D. Rennie, et al. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 354: 1896-900.

Moher, D., K. F. Schultz and D. G. Altman (2001). The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Paralell-Group Randomized Trials. *Annals of Internal Medicine* 134(8): 657-62.

Nichols, D. M., D. Pemberton, S. Dalhoumi, O. Larouk, C. Belisle, et al. (2000). DEBORA: developing an interface to support collaboration in a digital library. Research and Advanced Technology for Digital Libraries. 4th European Conference, ECDL 2000. Proceedings, Lisbon, Portugal, 239-48.

NLM (2002). MEDLINE Fact Sheet. Available at: www.nlm.nih.gov/pubs/factsheets/medline.html

Northup, D. E., M. Moore-West, B. Skipper and S. Teaf (1983). Characteristics of clinical information-searching: investigation using the critical incident technique. *Journal of Medical Education* 58: 873-81.

Olkin, I. (2001). Short Course on Meta-Analysis for Evidence-Based Medicine. Medical Decision Making, San Diego,

Petitti, D. B. (2000). Meta-Analysis, Decision Analysis and Cost-Effectiveness Analysis Methods for Quantitative Synthesis in Medicine. New York, Oxford University Press.

Petrosino, A. (1999). Lead Authors of Cochrane Reviews: Survey Results. Report to the Campbell Collaboration. Cambridge, MA, University of Pennsylvania.

Procter, R., A. Goldenberg, E. Davenport and A. McKinlay (1998). Genres in support of collaborative information retrieval in the virtual library. *Interacting with Computers* 10(2): 157-75.

Sim, I. (1998). Medical publishing meets AI. *IEEE Intelligent Systems* 13(1).

Sim, I., D. K. Owens, P. W. Lavori and G. D. Rennels (2001). A Complementary Method for Reporting Randomized Trials. *Medical Decision Making* 20(7): 440-50.

Smith, R. (1996). What clinical information do doctors need? *BMJ* 313(7064): 1062-1068.

Stroup, D. F., J. A. Berlin, S. C. Morton, I. Olkin, G. D. Williamson, et al. (2000). Meta-analysis of Observational Studies in Epidemiology. *JAMA* 283(15): 2008-12.

Sutton, A. J., K. R. Abrams, D. R. Jones, T. A. Sheldon and F. Song (1998). Systematic reviews of trials and other studies. *Health Technology Assess* 2(19).

Tengs, T., M. Adams, J. Pliskin, D. Safran, J. Siegel, et al. (1995). Five-hundred life-saving interventions and their cost-effectiveness. *Risk Analysis* 15(3): 369-390.

Tengs, T. and N. D. Osgood (2001). The link between smoking and Impotence : Two Decades of Evidence. *Preventive Medicine* 32(6): 447-452.

Twidale, M. and D. Nichols (1998). Designing interfaces to support collaboration in information retrieval. *Interacting with Computers* 10(2): 177-93.

Williamson, J. W., P. German, S., R. Weiss, E. A. Skinner and F. Bowes (1989). Health sciences information management and continuing education of physicians. *Annals of Internal Medicine* 110: 151-160.

Woolf, S. H. and D. A. Benson (1989). The medical information needs of internists and pediatricians at an academic medical center. *Bulletin of the Medical Library Association* 77(4): 372-80.