

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, they now must face the challenge of aggregating the information from within the retrieved documents. Our study is a first step toward both understanding this scientific integration process and developing tools to assist researchers with that process. We found that scientists actively collaborated as they refined the retrieval, extraction, and analysis phases of a process that we have called information synthesis. In this paper, we describe both our findings of their information behavior and our design and progress towards implementing a tool we call METIS (Multi-user ExTraction and Information Synthesis), which will support the collaborative, iterative, interactive information synthesis process used by public health and biomedical scientists.

1 Introduction

To answer research questions and make scientific advances, scientists often need to integrate information from the literature, but its current volume and growth rate impede their manual efforts. Scientists need automated tools to help them with the entire process of retrieving appropriate documents, extracting the needed data from within the documents, and analyzing the results. We refer to this three-phased process as **information syntheses**.

The focus of our research is both on understanding this information synthesis process and on developing new information technology to assist scientists in this process for the domain of public health and biomedicine. These scientists have a long tradition of using peer-reviewed literature in their quest to answer research questions. When studies differ in their conclusions or sample sizes are small, scientists use a rigorous non-biased methodology to collect all relevant articles, extract information from those articles

and integrate the results. The health and medical communities, refer to this process as a **systematic review**. A recent survey indicated that the time between an initial idea for a systematic review (one type of an information synthesis process) 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 on a risk factor (such as smoking) and breast cancer, 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 approaches will become even more difficult to use than they already are.

Our research explores ways to enable public health and biomedical scientists to use the medical literature faster and more completely than the current manual approaches allow. To gain the insights necessary to achieve this goal, we need to understand the context of the information synthesis process. Our study is a first step toward 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.

Accepted for the Proceedings of the 65th Annual Meeting of the American Society of Information Science and Technology (ASIST) Nov 2002.

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.

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 et al., 2000) in settings as diverse as rural clinics (Bowden et al., 1994) and academic medical centers (Woolf & Benson, 1989). Researchers have used various methods, such as observations (Forsythe, 1998), surveys (Williamson, et al, 1989), and interviews (usually based on the critical-incident technique) (Northup, 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. Our work examines this important user group for the task of information synthesis.

Collaboration is another aspect of information behavior that the medical area has largely neglected. Recently, the National Science Foundation funded a joint project with members from the University of Washington, Microsoft, Boeing, and Risø National Laboratory to explore such collaborative information seeking in engineering and software-development teams (Fidel et al., 2000). However, 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 & 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, et al, 1998) but few support teams collaborating to search for as well as use documents. A few exceptions are for systems that support collaborative browsing among many individuals (Nichols et al., 2000; Twidale & 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 authors must describe their rigorous, repeatable, and comprehensive search strategy. Their research question (and thus, their information need) is well defined, and is more appropriate for searching, rather than such browsing activities.

The synthesis process in general also can be viewed as a type of summarization of multiple documents. Several computer-science researchers have developed multi-document summarization systems (Mani & Bloedorn, 1999; McKeown & Radev, 1995). These systems use natural language processing techniques either to (1) extract representative sentences and connect them into a textual summary or (2) generate new sentences for the text summary based on the content of representative sentences. However, for information synthesis in medicine, the researchers care little about the natural-language sentences and paragraphs; they are most interested in extracting specific, quantitative and qualitative values from the text.

We are aware of one other approach that proposes to support information synthesis in a medical setting (Sim, et al, 2001). In her model, authors would submit the results of their clinical trial as both a database entry to a "trial bank" in addition to the published article. However, such an approach would not allow synthesis of material that was published before the trial bank, nor would it be applicable for the cases that will never be or have yet to be entered into the trial bank.

An important component of our approach is that we will semi-automatically extract from medical articles the information required by scientists for information synthesis. The Automatic Content Extraction (ACE) is a program within the National Institute of Standards and Technology that is developing technologies to extract information from news articles, speech and images (Doddington, 2000). Although research is underway to make these techniques more general (they are currently domain specific) (Pazienza, 1997, 1999) they do not consider (1) how information will be integrated after extraction; (2) the type of information required for information synthesis in public health and medicine or (3) the behavior of users who would use such information extraction tools.

3 Data Collection Methodology

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

3.1 Document Analysis

During our observations and interviews, we were actively reading material about the systematic review process. Our primary sources of information were documents produced by the Cochrane Collaboration (Clarke & Oxman, 2001) and **Health Technology Assessment (HTA)** (Sutton, Abrams, Jones, Sheldon, & Song, 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 & Schmidt, 1990; Ingelfinger, et al, 1994; Lipsey & Wilson, 2000; Petitti, 2000) and articles (Bartolucci, 1999; Davies & Crombie, 1998a, 1998b; DerSimonian & Laird, 1986; Engels, et al, 2000; Petrosino, 1999; Stroup et al., 2000).

Specifically, we were interested in how scientists use information from published literature in their work. We found recommendations from three working groups that described the type of information that should be included in three types 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 et al., 2000; Moher et al., 1999) and public health meta-analysis (Stroup et al., 2000) and the last group, CONSORT made recommendations for

information reported for randomized clinical trials (Begg et al., 1996), (Moher, et al., 2001). The documents that we read helped to provide us with insight into how the scientists would use this information in the scientific literature.

A recommendation of what should be included in a systematic review also provided us with insight into the kind of information to extract. To create the **extraction form** (a document that specifies what data should be extracted), scientists use either their own previous studies, similar studies from the literature, or external sources (such as the HTA or CC). We observed that extraction forms included a combination of general and specific information types. For example, a systematic review of the relationship between tobacco use and breast cancer would extract the subjects' age and location of the study in addition to how much they smoked. Exploring the relationship between alcohol and breast cancer would also require access to age and location information (discussed further in section 5.3)

3.2 Formal Training

The first author attended a presentation and short course to accelerate her familiarity with the systematic review process. The presentation was an open lecture at the University of California, conducted by Dr Adams on the systematic reviews. The short course was a half-day course on Meta Analysis that was conducted by Dr Olkin as part of a Decision Making Conference in San Diego (Olkin, 2001).

3.3 Observations

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 how we observed 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, group membership was established 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.

In addition to defining the research question clearly, the group also decided the type 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.

The CAM scientists identified studies to be included in their analysis using a multi-database, multi-search strategy. They then augmented the result by manually searching the references and by additional recommendations from experts. The experts in spinal manipulation suggested preliminary terms to the medical librarian, who then formulated the search strings for each database. The librarian in this study used eleven databases including MEDLINE, Ovid, CINAHL, and EmBase. For each search, the librarian e-mailed the citations details, including title and abstract to the entire group for review. The groups provided additional search terms and brought articles that were not identified by the database searches to the regularly held meetings. After the group was confident that they had retrieved the majority of relevant studies, they augmented the set by hand-searching the references of each article. They then sought recommendations by experts outside the group. The group provided each external expert with a complete bibliography and asked the expert to recommend additional literature, including gray literature. **Gray literature** is work that has yet to be 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 (Clarke & Oxman, 2001; Sutton et al., 1998). To verify that the extraction requirements were clearly stated, three group members and an expert in spinal manipulation (who had not been involved with the development of the extraction rules) used the rules to extract information from three articles, and then they compared their extractions with each other. Although this process verified that the extraction rules were clear (they all extracted the same information), the extraction methodology also needed to include a process for dealing with discrepancies. Thus, they also had three different scientists independently extract information from each article. If the resulting information was different, the three group members would explain their rationale, and if the discrepancy could not be resolved within the sub-group, a fourth person would consider the evidence and make the final decision.

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 et al., 1995). The first author observed their behavior by working in the same office for two days a week during summer 2001. Although we have 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 & 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).

The statistician, 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 result².

3.4 Interviews

During the preliminary meetings, members from both groups answered many questions regarding the process in general, the development of a search string, and the data-extraction 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

² The relationship between smoking and breast cancer is complex, with some reports suggesting that smoking is protective due to its effect on other factors, such as weight and menopause.

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

Scientists should make the methods that they used to identify, extract, and analyze information explicit, rigorous, non-biased, and repeatable. Although these traits are the cornerstone of systematic review of biomedical literature, we argue that they are true of good scientists in other disciplines. In addition to providing a coherent non-biased view of areas of research that are well studied, IS can be used to identify areas within medicine or public health where further research is required.

Both qualitative and quantitative systematic reviews can be categorized as using a 5-step process as follows: (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 & Crombie, 1998a). We have captured this process in figure 1, which illustrates that the research question is central to the IS process. Although the strategy for answering this question was refined, 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?*

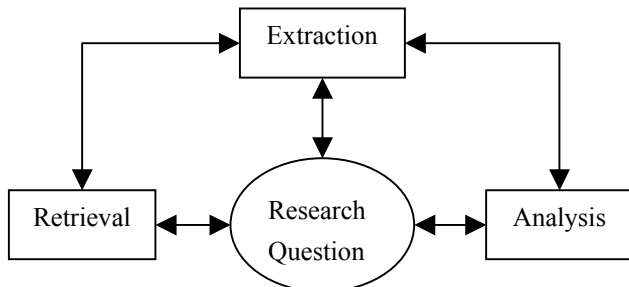


Figure 1 – The Information Synthesis Process. Scientists collect articles, extract information from those articles, and analyze the extracted information. The result of each phase enables scientists to refine the search criteria and extraction requirements. The availability of information influences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

During the retrieval phase, scientists require all potentially relevant articles; thus, recall is critical to a good systematic review. 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.

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 a clearly defined, search and inclusion criteria 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 with that information) is used to group articles together for analysis. For example, findings from low, medium and high quality studies are generally synthesized separately.

After scientists develop the data extraction form, they use a manual process to ensure that the information extracted is accurate. Three members of the medical group independently extracted information from each article and then compared their results. In the public health group, information was extracted by one person and stored in a database. To ensure accuracy, two other group members independently verified the extracted information. We observed that people extracting information from the documents used the context of the document. Authors adjust information within a table of an article, such as adjusting for age; thus, the person extracting and verifying information requires access to the document content. This process ensures 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 & Laird, 1986) after pre-processing both the information extracted from articles data from the BRFSS.

5 Functional Requirements

The user behavior that we observed included how group members interacted with other members, the information synthesis task and with other groups. We observed that

each of these interactions required collaboration because each scientist brought to the group a unique, but complementary set of skills. We also observed that the group iterated through each of the retrieval, extraction and analysis phases of IS. A system that does not offer support for these user behaviors will not be successful in this domain.

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 sections 5.3.

5.1 Collaboration

The first user behavior that we observed was that group members actively collaborated throughout the information synthesis process. This 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. We collected all meta-analyses in MEDLINE between 1990 and 1996 that were published in the top medical journals³. Of the 147 studies that satisfied this criterion, 132 had multiple authors. This indicates that a systematic review of biomedical literature is often a collaborative exercise. This user community does appear to have an affinity for collaborative information exchange. This is apparent in a clinical setting where physicians depend on each other (rather than the medical literature) to satisfy their information needs related to patient care (Ely, 1999) (Smith, 1996).

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 the framework proposed by Davis and Crombie (Davies & Crombie, 1998a) to describe the collaborative and individual activity involved in information synthesis.

(1) Defining the research question

The process of defining the research question was highly collaborative. Users held face-to-face meetings to discuss and refine candidate research questions. Group members from the medical group often brought 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) Searching literature

In the medical group, the librarian was central to collaboration during the literature search. She used her expertise to formulate a search string for each database based on descriptions that were by experts in spinal manipulation. The choice of databases searched was a collaborative decision with input from all group members. The medical librarian provided the title and abstract of each search result to the other group members who then identified additional terms.

Recall that three group members read each article. The first group member to read the article hand searches the references to provide the medical librarian with additional citations. The librarian then obtains and distributes the new article to the group member. In this way, the expertise of domain expert is coupled with the retrieval expertise of the librarian, to ensure that all relevant articles are collected.

In the public health group, the primary investigator and a research assistant collaborated on developing a search string that will ensure that all relevant articles are retrieved. Unlike the area of complementary and alternative medicine, MEDLINE contains much of literature in public health, thus this was the only database used to suggest articles. The full text of each citation from MEDLINE was obtained from the University libraries.

We also observed a passive collaboration between the group and previous studies. Specifically authors of a systematic review make their search strategy explicit in their published manuscript. In this way, the authors of a previously published systematic review are referenced during the development of a new search strategy. The medical group also consulted recommendations from the HTA and CC.

(3) Assessing studies

The extracted information and importance score to each information item determined the quality of each article. 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 were 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.

³ Based on a previous empirical analysis of meta-analysis (Engels 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.

(4) Combining results

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). If a qualitative approach is adopted, the studies are generally presented and discussed with respect to the quality of the article. Although the biostatistician played a key 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.

(5) Placing the findings in context

Both groups collaborated during the development of the manuscript. 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 needs information requirements during each of the retrieval, extraction and analysis phases in the IS process. Scientists documented this evolution, particularly the search strategy, because they are required to provide these details when communicating the results of the synthesis to other scientists.

The artifact that reflected this refinement during retrieval was a 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 the additional articles obtained using these terms and repeat the process. The librarian maintained a citation database to manage information regarding the articles considered from database searches, by hand-searching the references of each article, and from recommendations made by experts outside the group.

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. In their case, the inclusion criterion could have been used to constrain the articles that were considered, however such a refinement would only be possible if the full text was available. Specifically they 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.

5.3 Information Requirements

In addition to supporting user behavior, we must also provide the facts that public health and biomedical scientists use in the information synthesis process. We found that the information required for synthesis is often only located within an article, not in the title and abstract of an article. This observation has implications to indexing, retrieval and extraction strategies. If search engine developers do not use full text when generating indexes, a user will be unable to specify a high precision query. Consider again, the public health group who required all impotence studies that reported tobacco usage. We observed that authors report tobacco usage, even when this factor did not play a major role in their study. Authors would therefore refer to impotence (or a synonym) in the title, abstract or keywords of a relevant article, however authors would only report tobacco usage and other low impact factors within the article. For a study to be valid, a necessary (but not sufficient) criterion is that the retrieval strategy must achieve close to perfect recall. Thus, it was necessary to review manually all articles on impotence to identify those that reported smoking habits.

The location of information has important consequences to the development of extraction algorithms. As in the retrieval phase, the full text of an article is required for extraction. 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 of participants

and tobacco usage. This suggests extraction algorithms can use the structure of a document to identify certain facts.

We were surprised that certain information extracted for the analysis of impotence and smoking was 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. While this list is certainly not comprehensive, we include it to demonstrate that regularities in the type of information required for information synthesis exist. We ordered the categories in increasing specificity to a review.

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 (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 can also be used during face-to-face meetings of group members, such as formulating the research question. 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.

Currently, scientists record the history of this search string development, because it is critical that the final search string be included in the systematic review. METIS will follow the trend of information retrieval systems that

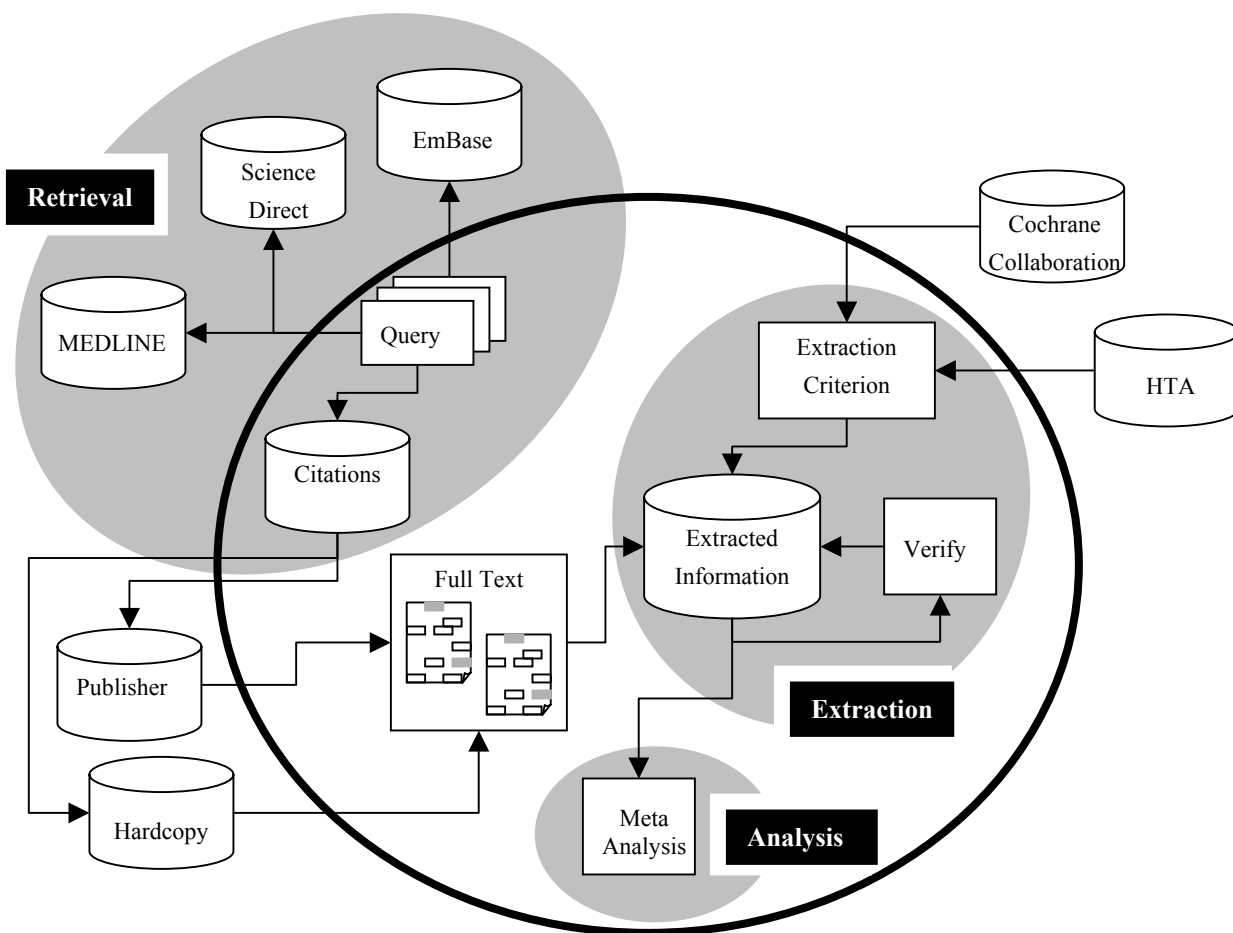


Figure 3 – METIS System Architecture. METIS is an integrated retrieval and extraction system that enables scientists to perform a systematic review of biomedical literature more effectively than the current manual approaches. METIS extracts information from within the full text of articles, and then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

provide the functionality to save this information, such as Wiley Interscience that enables users to save multiple queries. We anticipate that scientists might want to search multiple databases. Thus, our future version 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 available from the CC to retrieve only those articles from MEDLINE that are suitable for a medical meta-analysis should be simple (Clarke & 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 is 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, such as the Alembic Workbench and HPIEW that automatically create rules to extract information from text (Caruana et al., 2000; Day et al., 1997). We are currently reviewing the interface of this component with experts in breast cancer.

6.4 Analysis

Scientists refined their retrieval and extraction strategies after they had conducted preliminary analyses, such as the number of articles that reported a particular fact. By integrating analysis functionality with the retrieval and extraction components, we hope to reduce the time taken to revise each phase and thus enable users to explore other types of connections.

METIS will (1) provide a summary of all the information extracted from all articles; (2) enable a scientist to pre-process extracted information, before conducting a meta-analysis, and (3) conduct and visualize the results of a meta-analysis. Scientists must decide the appropriate scaling for meta-analysis, such as using an odds ratio or relative risk metric. We have yet to implement (1) and (2). We have completed a METIS component (in Java) to perform (3) a random effects meta-analysis (DerSimonian & 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, McLennan, & Rothman, 2001). We are confident that this module works because our results for smoking and impotence meta-analysis are the same as those reported in the smoking and impotence study by the public health group (Tengs & Osgood, 2001).

7 Conclusion

The goal of our study was twofold: (1) to gain an understanding of how scientists in public health and medicine use biomedical literature to answer research questions and (2) to design a system that would support this behavior.

Our main findings were that retrieval is part of a larger process, which we call information synthesis, and that scientists iterate between the retrieval, extraction, and analysis phases of this process. We also found that although scientists have a clear, well-defined information need, this criteria alone was insufficient to escape the problem of information overload. To support the user behavior that we observed and thus reduce information overload, a system

also needs to support the extraction and analysis phases of the information synthesis process.

We found that users actively collaborated, and our design of the METIS system reflects this finding. Other findings that heavily influenced our system design were that scientists use both automated and manual mechanisms to (1) perform a comprehensive search of documents related to the research question and, (2) to verify the accuracy of the information extracted. These findings imply that public health and medical scientists would not accept a completely automated approach to information synthesis.

This paper reports a first step toward understanding the information synthesis process used by public health and biomedical scientists. By working with scientists in the context of their work, we were able to gain an understanding of their information requirements, and were then able to use this knowledge to design a system that would support their information behavior. 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 Samuelli 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

- Asilomar 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
- Allais, G., Voghera, D., DeLorenzo, C., Mana, O., & Benedetto, C. (2000). Access to Databases in Complementary Medicine. *The Journal of Alternative and Complementary Medicine*, 6(3), 265-274.
- Bartolucci, A. (1999). The Significance of Clinical Trials and the Role of Meta-Analyses. *Journal of Surgical Oncology*, 72, 121-3.
- Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K. F., Simel, D., & Stroup, D. F. (1996). Improving the Quality of Reporting of Randomized Controlled Trials: The CONSORT Statement. *JAMA*, 276, 637-9.
- Bowden, V. M., Kromer, M. E., & Tobia, R. C. (1994). Assessment of physicians' information needs in five Texas counties. *Bull Med Libr Assoc*, 82(2), 189-196.
- BRFSS. (2001). Behavioral Risk Factor Surveillance System, NCCDP and Health Promotion, CDCP, U.S. <http://www.cdc.gov/nccdphp/brfss/>.
- Caruana, R., Hodor, P. G., & Rosenberg, J. (2000). *High Precision Information Extraction*. Paper presented at the KDD-2000 Workshop on Text Mining.
- Clarke, M., & Oxman, A. D. (2001). Cochrane Reviewers' Handbook 4.1.1 [updated October 2001], *The Cochrane Library*. Oxford.
- Davies, H. T. O., & Crombie, I. K. (1998a). What is a systematic review? *Hayward Medical Communications*.
- Davies, H. T. O., & Crombie, I. K. (1998b). What is meta-analysis? *Hayward Medical Communications*.
- Day, D., Aberdeen, J., Hirschman, L., Kozierek, R., Robinson, P., & Vilain, M. (1997, 31 March -- 3 April). *Mixed-Initiative Development of Language Processing Systems*. Paper presented at the Fifth Conference on Applied Natural Language Processing, Washington.
- DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177-88.
- Doddington, G. R. (2000) ACE Automatic Content Extraction. Presentation at the TIDES program Kick-off meeting. <http://www.nist.gov/speech/tests/ace/index.htm>
- Ehrlich, K., & Cash, D. (1994). *Turning Information into Knowledge: Information Finding as a Collaborative Activity*. Paper presented at the Digital Libraries '94.
- Ellison, R. C., Zhang, Y., McLennan, C. E., & Rothman, K. J. (2001). Exploring the Relation of Alcohol Consumption to Risk of Breast Cancer. *American Journal of Epidemiology*, 154(8), 740-7.
- Ely, J. W., Osheroff, J. A., Ebell, M. H., Bergus, G. R., Levy, B. T., Chambliss, M. L., Evans, E. R. (1999). Analysis of questions asked by family doctors regarding patient care. *BMJ*, 319, 358-361.
- Engels, E. A., Schmid, C. H., Terrin, N., Olkin, I., & Lau, J. (2000). Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Statistics In Medicine*, 19(13), 1707-28.
- Fidel, R., Bruce, H., Pejtersen, A. M., Dumais, S., Grudin, J., & Poltrock, S. (2000, August 16-18, 2000). *Collaborative Information Retrieval (CIR)*. Paper presented at the Information Seeking in Context 2000: The 3rd International Conference on Information Seeking and Use in Different Contexts, Goteborg, Sweden.
- Forsythe, D. (1998). Using ethnography to investigate life scientists' information needs. *Bulletin of Medical Library Association*, 86(3), 402-9.
- Fowell, S., & Levy, P. (1995). Developing a New Professional Practice: A Model for Networked Learner Support in Higher Education. *Journal of Documentation*, 51(3), 271-80.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: correcting error and bias in research findings*. Newbury Park, CA: Sage.

- Ingelfinger, J. A., Mosteller, F., Thibodeau, L. A., & Ware, J. H. (1994). *Biostatistics in clinical medicine* (3rd ed.): McGraw-Hill Inc.
- Lange, L. L. (1993). *Information Seeking by Nurses During Beginning-of-Shift Activities*. Paper presented at the Proceedings of the AMIA Fall Symposium.
- Leydon, G. M., Boulton, M., Clare Moynihan, Jones, A., Mossman, J., Boudioni, M., & McPherson, K. (2000). Cancer patients' information needs and information seeking behaviour: in depth interview study. *British Medical Journal (BMJ)*, 320, 909-13.
- Lipsey, M. W., & Wilson, D. B. (2000). *Practical Meta-Analysis* (Vol. 49): Sage.
- Mani, I., & Bloedorn, E. (1999). Summarizing Similarities and Differences Among Related Documents. In I. Mani & M. T. Maybury (Eds.), *Advances in Automatic Text Summarization* (pp. 357-79): The MIT Press.
- McKeown, K., & Radev, D. R. (1995). *Generating Summaries of Multiple News Articles*. Paper presented at the Proceedings, 18th Annual International Conference on Research and Development in Information Retrieval.
- Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., & Stroup, D. F. (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., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., Stroup, D. F., & Group, Q. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet*, 354, 1896-900.
- Moher, D., Schultz, K. F., & Altman, D. G. (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., Pemberton, D., Dalhoumi, S., Larouk, O., Belisle, C., Twidale, M. B., Borbinha, J., & Baker, T. (2000, Sept. 2000). *DEBORA: developing an interface to support collaboration in a digital library*. Paper presented at the Research and Advanced Technology for Digital Libraries. 4th ECDL 2000. Proceedings (Lecture Notes in Computer Science Vol.1923), Lisbon, Portugal.
- NLM. (2002, 17 September 2001). *MEDLINE Fact Sheet*. National Library of Medicine. Available: www.nlm.nih.gov/pubs/factsheets/medline.html.
- Northup, D. E., Moore-West, M., Skipper, B., & Teaf, S. (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*. Paper presented at the Medical Decision Making, San Diego.
- Pazienza, M. T. (1997). *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology; International Summer School SCIE-97* (Vol. 1299). Frascati, Italy: Springer.
- Pazienza, M. T. (1999). *Information Extraction: Towards Scalable, Adaptable Systems* (Vol. 1714). Frascati, Italy: Springer.
- Petitti, D. B. (2000). *Meta-Analysis, Decision Analysis and Cost-Effectiveness Analysis Methods for Quantitative Synthesis in Medicine* (2nd ed. Vol. 31). 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., Goldenberg, A., Davenport, E., & McKinlay, A. (1998). Genres in support of collaborative information retrieval in the virtual library. *Interacting with Computers*, 10(2), 157-75.
- Sim, I., Owens, D. K., Lavori, P. W., & Rennels, G. D. (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), 1062-8.
- Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., Moher, D., Becker, B. J., Sipe, T. A., & Thacker, S. B. (2000). Meta-analysis of Observational Studies in Epidemiology. *JAMA*, 283(15), 2008-12.
- Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A., & Song, F. (1998). Systematic reviews of trials and other studies. *Health Technology Assess*, 2(19).
- Tengs, T., Adams, M., Pliskin, J., Safran, D., Siegel, J., Weinstein, M., & Graham, J. (1995). Five-hundred life-saving interventions and their cost-effectiveness. *Risk Analysis*, 15(3), 369-90.
- Tengs, T., & Osgood, N. D. (2001). The link between smoking and Impotence : Two Decades of Evidence. *Preventive Medicine*, 32(6), 447-52.
- Twidale, M., & Nichols, D. (1998). Designing interfaces to support collaboration in information retrieval. *Interacting with Computers*, 10(2), 177-93.
- Williamson, J. W., German, P., S., Weiss, R., Skinner, E. A., & Bowes, F. (1989). Health sciences information management and continuing education of physicians. *Annals of Internal Medicine*, 110, 151-60.
- Woolf, S. H., & Benson, D. A. (1989). The medical information needs of internists and pediatricians at an academic medical center. *Bulletin of the Medical Library Association*, 77(4), 372-80.