Collaborative Information Synthesis I: A Model of Information Behaviors of Scientists in Medicine and Public Health

Catherine Blake
School of Information and Library Science, University of North Carolina, Chapel Hill, NC 27599. E-mail: cablake@email.unc.edu

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

Scientists engage in the discovery process more than any other user population, yet their day-to-day activities are often elusive. One activity that consumes much of a scientist's time is developing models that balance contradictory and redundant evidence. Driven by our desire to understand the information behaviors of this important user group, and the behaviors of scientific discovery in general, we conducted an observational study of academic research scientists as they resolved different experimental results reported in the biomedical literature. This article is the first of two that reports our findings. In this article, we introduce the Collaborative Information Synthesis (CIS) model that reflects the salient information behaviors that we observed. The CIS model emerges from a rich collection of qualitative data including interviews, electronic recordings of meetings, meeting minutes, e-mail communications, and extraction worksheets. Our findings suggest that scientists provide two information constructs: a hypothesis projection and context information. They also engage in four critical tasks: retrieval, extraction, verification, and analysis. The findings also suggest that science is not an individual but rather a collaborative activity and that scientists use the results of one analysis to inform new analyses. In Part 2, we compare and contrast existing information and cognitive models that have inadvertently reported synthesis, and then provide five recommendations that will enable designers to build information systems that support the important synthesis activity.

Introduction

Scientists engage in the discovery process more than any other user population, yet their day-to-day activities are often elusive. Even a scientist who actively makes discoveries in one discipline can find the activities conducted in a related field a mystery. Regardless of their specific discipline, the role of a good scientist is to develop a model of the world that accurately explains the available evidence. The development of accurate models often requires that a scientist resolve conflicting evidence.

One activity that consumes much of a scientists' time is *synthesis*, "the dialectic combination of thesis and antithesis into a higher stage of truth" (*Merriam-Webster's Collegiate Dictionary*, 2004). This dictionary definition reflects the alternative viewpoints that often occur when multiple empirical studies explore the same phenomena. The synthesis activity results in an overall finding—a higher stage of truth—which scientists achieve by resolving conflicting evidence. Thus, the synthesis activity requires accurately weighing a body of evidence that includes *contradictions* (when the study results differ) and *redundancies* (when study results concur) that are inevitable when multiple studies explore the same natural phenomena. In this article, we consider synthesis activities that involve evidence reported in existing literature rather than synthesis activities that require additional data collection through experimentation.

New technology and changes in publishing practices continue to increase the quantity of literature available to scientists. For example, a breast cancer scientist already can access abstracts of more than 122,000[1] articles, and by this time next year, he or she will have access to an additional 5,400 new abstracts.[2] Scientists could reduce the number of

Received October 14, 2004; revised August 12, 2005; accepted October 30, 2005

 • Published online 31 August 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20487

[1]Search results from a query conducted on July 20, 2005, in pubmed.org using the medical subject heading "breast neoplasms" identified 122,560 citations.

[2]Average number of articles with the medical subject heading "breast neoplasms" per year over the last decade.

articles they read by reading reviews produced by other scientists; however, review articles are not available for every topic area, and articles that do not include recently obtained evidence are of little value. When a relevant, current review is not available, scientists face the daunting task of integrating evidence from hundreds or thousands of empirical studies reported in the literature.

Integrating information from biomedical literature is a time-consuming task. One survey of 37 groups of scientists revealed a total mean time of 1,139 hours (Allen & Olkin, 1999). Assuming that our scientist could dedicate 8 hours a day exclusively to the synthesis activity, it would take approximately 7 months. Another survey of 14 scientists revealed an elapsed time of 28 months between an initial review idea and its later publication (Petrosino, 1999). Our scientist could reduce this effort by constraining the scope; however, this reduction could introduce undesirable biases and thus reduce the validity of the entire synthesis activity.

Although synthesis is just one activity a scientist conducts, the time required and the potentially central role in discovery warrant detailed investigation. This article is the first of two that reports our findings. In Part 1, we introduce the *collaborative information synthesis* model that captures the salient information behaviors that we observed. Our findings suggest that during the synthesis activities, scientists provide two information constructs (the hypothesis projection and context information) and engage in four critical tasks (retrieval, extraction, verification, and analysis). Our results suggest that synthesis is a collaborative rather than an individual activity and that a scientist will iterate both within and between critical tasks. We also found that scientists initiate additional synthesis activities based on previous analyses. In Part 2 (Blake & Pratt, in press), we compare and contrast the information and cognitive science models that have inadvertently identified synthesis behaviors and provide additional details regarding the medical and public health groups that we considered, and specific system design recommendations that we believe will best support synthesis activities.

Background

In this section, we summarize the systematic review process and the study environment, which provides the background necessary to interpret our findings and understand the challenges faced by scientists during synthesis activities. We also outline the mixed methods approach that we used to collect data for this study. In the interests of conserving space in this article, we report additional details of the user groups in a second article regarding this study (Blake & Pratt, in press).

The Systematic Review Process

Two organizations play an active role in establishing the methodology used to conduct a systematic review. The Cochrane Collaboration (www.cochrane.org) is a collection of experts who, on a voluntary basis, provide both methodological advice (Alderson, Green, & Higgins, 2004) and access to systematic reviews. The Health Technology Assessment Program (http://www.hta.nhsweb.nhs.uk/) also provides methodological guidance (Sutton, Abrams, Jones, Sheldon, & Song, 1998). Both organizations describe a five-stage systematic review process that includes the following: (1) Define a research question, (2) search the literature, (3) assess study quality, (4) combine findings, and (5) place the findings in context. Although each of these steps are important to the systematic review process, Steps 3 to 5 are of most interest in this study because they provide insight into the synthesis activities employed by scientists in medicine and public health.

From an information science perspective, Bates (1976) also emphasized the importance of clearly defining the inclusion and exclusion criterion of a rigorous systematic bibliography. The quality scores assigned during Step 3 correspond with the differentiating stage in Ellis's (1989, 1993; Ellis & Haugan, 1997) existing models of information behaviors.

Study Environment

To understand the application of the systematic review process, we conducted our study in two naturalistic settings: medicine and public health. The medical group formed specifically to conduct nonbiased, rigorous reviews of the literature relating to complementary and alternative approaches to medicine. The members ($n = 8$) of the medical group had expertise in systematic review methodologies, library science, biostatistics, health services, clinical research, and clinical content. We attended their first organizational meeting of the medical group on July 10, 2001. During the first meeting, the medical group decided to focus future meetings on one of three topics: the organization of their review, the search strategy, or methodological issues. Based on the discussion at that meeting, the first author decided that the methodology meetings would provide the most insight into the general synthesis behaviors. Thus, the first author attended, observed directly, and recorded the methodological meetings of the medical group between July and September 2001.

In contrast to the medical group, the participants ($n = 11$) in the public health setting (i.e., the public health group) had worked together for several years before our study began. Students were the only exception to this rule: Each graduate student had worked in the group for less than 1 year, and the majority of undergraduate interns had worked in the group for less than 1 month. The first author worked from the public health group's offices for 2 days a week in Summer 2001 to observe their synthesis activities. During that time, three synthesis activities were in progress: (a) a systematic review that explored the relationship between smoking and impotence, (b) a meta-analysis on utility estimates and AIDS, and (c) an ongoing project that centers on the creation and maintenance of a database comprising lifesaving and cost-effectiveness data. Each project was in a different stage of

DOI: 10.1002/asi

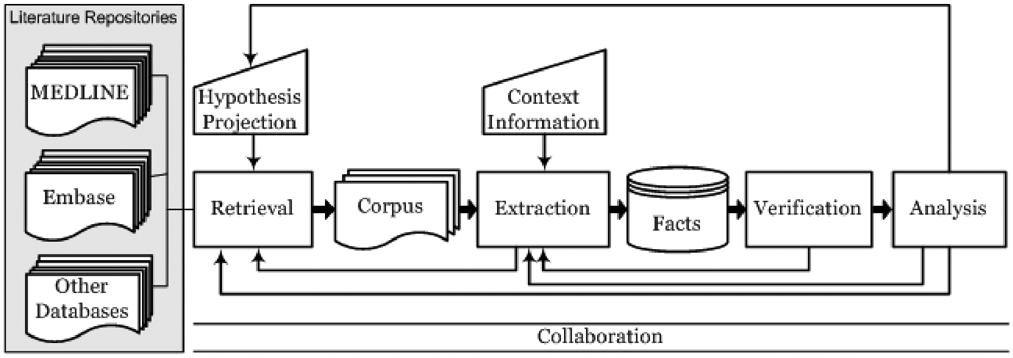

FIG. 1. The Collaborative Information Synthesis (CIS) Model.

completion and comprised a different subset of the total public health group, which enabled us to triangulate their information behaviors better than if we had studied one project in isolation.

Data-Collection Methodology

The collaborative information synthesis model emerged from a rich collection of qualitative data collected from both the medical and public health groups. We provide a complete description of the methodology in Blake and Pratt (in press). The first author observed the methodological meetings of the medical group as they conducted the systematic review. In addition to the direct observations, we asked short clarification questions that addressed the medical group's search criteria and analysis methodologies and interviewed the domain expert using an open-ended question format. We supplemented the direct observations and interviews with information artifacts including (a) minutes from methodology meetings produced by the group as part of their process; (b) minutes from search strategy meetings, also produced by the group; (c) bibliographic references collected during the group's initial search; (d) methodology literature recommended by the group; and (e) literature in the library and online.

In contrast to the solely prospective nature of our data-collection methods with the medical group, our methods for the public health group were both prospective and retrospective because the projects had begun prior to the initiation of our study. Data from the public health group comprised interviews, observations, and information artifacts. We conducted an open-ended, face-to-face interview with the director and a separate open-ended, face-to-face interview with the director and the statistician. The research programmer–statistician had left the group, so we interviewed him by telephone, e-mail, and had a short, face-to-face interview. The first author worked from the public health group office 2 days a week during Summer 2001. During that time, she observed and conducted open-ended discussions with interns who were developing the group's cost-effectiveness database.

Findings and Discussion

Two kinds of information constructs emerged from our rich collection of qualitative data: a hypothesis projection and context information. Those constructs intersected with the four critical tasks: retrieval, extraction, verification, and analysis. In addition to those critical tasks, we identified two information behaviors that spanned both the information constructs and the critical tasks, which we have labeled *iteration* and *collaboration*. We have called our model *collaborative information synthesis* (see Figure 1) to reflect the importance of both collaboration and synthesis that we consider underemphasized in existing models of information behavior. In this section, we describe each of our characterizations and provide specific examples from the extensive qualitative data collected.

Information Constructs Provided by Scientific Users

The qualitative data revealed two kinds of user-provided information constructs: a hypothesis projection and context information. In this section, we describe each of these constructs. In addition to these constructs, both user groups had access to collections of reference databases. The medical group used 10 databases because they were not convinced that MEDLINE sufficiently covered articles on spinal manipulation. In contrast, the public health group used only the MEDLINE database, the most widely used reference database in life sciences.

Hypothesis Projection

Our literature review of the systematic review process indicated that they start with a "problem specification" (Lipsey & Wilson, 2000) or "an appropriate therapeutic question" (Davies & Crombie, 1998) In contrast to a specific research question, the question posed by experts in both the medical and public health groups is best described as a hypothesis projection. According to Rescher (1978), hypothesis projection is

> the purely conjectural proliferation of a whole gamut of alternative explanatory hypotheses that are relatively plausible, a proliferation based on guesswork—though not 'mere' guesswork, but guesswork guided by a scientifically trained intuition. The aim of this enterprise is to identify those hypotheses that merit detailed scrutiny. (p. 8)

In contrast with the collection of potential hypotheses within the hypothesis projection, hypothesis testing is "the

DOI: 10.1002/asi

elimination of a hypothesis on the basis of observational data, generally secured by suitably contrived experimental trials" (Rescher, 1978, p. 8). In a health care setting, a hypothesis projection would be a set of candidate treatments, and hypothesis testing would require a traditional study such as a randomized clinical trial to evaluate the efficacy of each treatment.

The medical group's hypothesis projection explored spinal manipulation; however, their specific research question took many discussions and required an extensive search of existing literature to identify the kinds of studies that were available. The group explored several directions, including the reliability of spinal palpatory exams, the effectiveness of osteopathic spinal manipulation for low back pain, and the validity of spinal palpatory procedure for screening and diagnosis of patients with spinal neuromuscular dysfunction. Table 1 captures the evolution of the medical group's hypothesis projection and their final research question, which they explored in their systematic review.

The public health group also described an evolutionary process to develop their topic. During the initial phases of their hypothesis projection, they had considered including both smoking and alcohol consumption in their analysis. Although sufficient articles reported both smoking and alcohol consumption rates separately, the lack of articles that reported the joint probability of both smoking and alcohol consumption forced the public health group to restrict their hypothesis projection to only smoking.

A scientific user initiates the collaborative information-synthesis activity with a hypothesis projection. This starting point motivates the question "Where does the first hypothesis projection originate?" Although we did not directly observe either user group during the formation of the projects that we report here, we can comment on the source of the hypothesis projection. During the formation of the medical group's hypothesis projection for this study, they developed a variety of new hypotheses. They saved one of these hypothesis projections—the reliability of chiropractic procedures—as a future study. Thus, the group would create a new instantiation of the collaborative information synthesis activity to support their new reliability hypothesis projection.

TABLE 1. Evolution of the medical group's hypothesis projection.

July 10, 2001 (First organizational meeting)	
	A Systematic Review of the Effectiveness of Osteopathic Spinal Manipulation for low back pain
July 23, 2001	
	What is the reliability and validity of spinal palpatory diagnosis in diagnosis with spinal neuromuscular dysfunctions?
August 3, 2001	
	What is the reliability of spinal palpatory procedure(s)?
	What is the validity of spinal palpatory procedure for screening and diagnosis of patients with spinal neuromuscular dysfunctions?
May 7, 2003 (Published)	
	Content validity of manual spinal palpatory exams—A systematic review

From an information-retrieval perspective, the hypothesis projection results in a variety of search terms. In the medical group, the medical librarian tailored the search terms to each bibliographical database. Each collection of search terms reflected the group's current hypothesis projection. The retrieval-task section of this article describes in more detail the search methods used by the medical group.

Context Information

The medical group drew on four resources to identify specific information items that they should extract. The first three resources were external to the group and comprised previously published meta-analyses on a similar topic, sample extraction worksheets from the Cochrane Collaboration, and sample descriptions from the Health Technology Assessment group. The fourth and most challenging resource was the clinical experience possessed by the domain experts in the medical group. Defining the required information consumed the majority of the methodology meetings. Although the medical group used e-mail to clarify information items, we observed that the face-to-face interactions, including the use of a white board, were critical during the development of context information.

Our reviews of the methodology meeting notes revealed an interesting relationship between the medical group members who possessed clinical expertise and the members who were not domain experts. Although the information proposed by the domain experts was critical, group members who were experts in other fields also provided an important role in the project. The presence of non-domain-expert members required the domain experts to articulate assumptions that years of experience had firmly engrained into their mental model. This behavior is consistent with studies showing that clinical reasoning becomes more implicit as expertise increases (Elstein, Shulman, & Sprafka, 1978).

These interactions revealed both the challenge of collaborative knowledge-work and the benefit that can otherwise be difficult to obtain. It also suggests that extracting information from a user's cognitive model can be challenging.

After multiple iterations and long discussions, the medical group captured the context information on an extraction worksheet. An extraction worksheet is a working document within the systematic review process that captures the information required from each article that satisfies the inclusion criteria. Before the extraction task began, the medical group verified that they had sufficiently described the required context information on the extraction worksheet by conducting a pilot study. In their pilot study, three members from the medical group and an independent domain expert (who had not previously been involved in the project) used the extraction worksheets to identify facts from three articles. Each of the 4 test participants extracted similar information from the articles, supporting the clarity of the extraction rules.

Developing context information is a time-consuming process, and clearly articulating an exhaustive list is intractable; however, despite different hypothesis projections,

DOI: 10.1002/asi

we observed regularities in the information items required by both the medical and public health groups.

In addition to the hypothesis projection, our data revealed that scientists provide context information. Context information is the collection of facts that the group identified from each article, which could potentially influence the review outcomes. For example, context information from the public health group's study on smoking and impotence comprised (a) the total number of impotent subjects; (b) the number of impotent men who were current smokers; (c) the definition of impotence used in the study; (d) the definition of smoking used in the study; (e) mean, standard deviation, and range of ages in the subject population; (f) the geographic location of the study; (g) the time period over which assessment occurred; and (h) whether the article mentioned smoking in the abstract (Tengs & Osgood, 2001). Although not required by the formal definition of the systematic review process, our information artifacts showed that scientists often include the context information in their published article.

A closer inspection of data from the medical and public health groups and extraction worksheets from the Cochrane Collaboration and the Health Technology Assessment groups revealed a collection of typical context information, as shown in Table 2. The hypothesis projection couples loosely to study- and population-context information (i.e., many hypothesis projections require the same study and population information). In contrast, the hypothesis projection couples tightly to risk factor or intervention and the medical condition (i.e., the hypothesis projection strongly influences the type of risk factor, intervention, and medical information required). This coupling is shown in Table 2. The context information within Items 1 and 2 generalize to a variety of studies in medicine and public health while the context information within Items 3 and 4 are specific to a hypothesis projection.

TABLE 2. Example of the information required for Collaborative Information Synthesis.

1. **Information related to the study**
   - Number of subjects with medical condition (e.g., number of patients who are impotent)
   - Year of publication and year in which data were collected
   - 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
3. **Information related to the intervention or risk factor**
   - Details of the intervention or risk factor [e.g., kind of palpitation or type of tobacco (cigarettes, pipe)]
   - Amount of exposure to the intervention or risk factor (e.g., number of palpitations, 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)

Despite their different hypotheses projections, four categories of context information emerged from the studies conducted by the medical group and each public health group (see Table 2). Our characterization of context information is consistent with the standards developed by several medical groups, including (a) the checklist for clinical trials that was developed by the Asilomar Working Group (1996); (b) the guideline for randomized clinical trials (Begg et al., 1996); (c) the standards for meta-analyses based on observational studies (Stroup et al., 2000), which were developed by the MOOSE group; and (d) the standard for meta-analyses based on randomized clinical trials (Moher, Schulz, & Altman, 2001), which was developed by the QUOROM group. Our characterization also was consistent with other study descriptors (Lipsey & Wilson, 2000). Thus, despite differing hypotheses projections, context information appears to generalize between studies. The existence of similar context information suggests that a system that automatically extracts context information would enable scientists to explore a variety of hypothesis projections.

In addition to regularities in the nature of information required, the location of the information within an article did not differ between the two groups. Our analysis of information artifacts revealed that both groups required the full text of an article to collect all of their required context information. We also observed that information in tables played an important role in the synthesis activity.

Critical Tasks

Our data reveal that during the synthesis activity, scientists engage in four critical tasks: retrieval, extraction, verification, and analysis. In this section, we describe each task and provide examples from both the medical and public health groups.

Retrieval. The first critical task that emerged from our data was retrieval. The importance of a comprehensive literature search in a systematic review required that each group invest considerable effort to ensure that they had obtained a comprehensive collection of articles. Thus, recall was more important to this user population than was precision.

MEDLINE is the most frequently used bibliographic index (Hopewell, Clarke, Lefebvre, & Scherer, 2002); however, the inter-indexer reliability of the Medical Subject Headings (MeSH) assigned to each article and difficulty in matching MeSH terms to a user's information need make accurate retrieval problematic. For example, Funk and Reid (1983) reported that indexer agreement between the main MeSH assignments was only 61%, and agreement on subheadings was an astonishingly low 49%. Lack of consistency in assigning keywords adds to a user's retrieval challenge. Our data reveal that scientists are aware of these indexing inadequacies and that they use a multiple-strategy approach to overcome existing retrieval limitations. This behavior has been reported in other studies as well (Murphy et al., 2003a, 2003b). The medical group used five search strategies to ensure that they had retrieved a comprehensive collection of articles. The public health group stated during the interviews that they also developed their search strategy using an iterative process. We focus on the medical group in this discussion because we directly observed their process. The medical group conducted multiple searches, used multiple bibliographic databases, hand-searched important journals and references from each article, and sought additional contributions from experts in the field.

The medical group's first strategy to ensure a comprehensive article collection was to conduct multiple searches. The domain experts and the medical librarian developed the collection of initial search terms. Each week, the group added new search terms that they had identified from their previously collected articles or that had emerged during their face-to-face and e-mail discussions. The public health group also refined their search string throughout the review. These findings are consistent with the Cochrane Collaboration's recommendation to conduct multiple searches (Clarke, 2002). From an information science perspective, this result corresponds to the subject search identified by Bates' (1989) berry-picking model.

The medical group's second strategy was to search multiple bibliographic databases. The group tailored search strings to 10 separate bibliographic databases: MEDLINE, Ovid Mantis, CINAHL, EmBase, Web of Science, OCLC PaperFirst, Biosis Preview, Index to Chiropractic Literature, PEDro, Cochrane Library, and MDConsult. For each database, the librarian e-mailed citation details returned by the search strings, including the title and abstract, to the entire group for review.

The third strategy was to hand-search important journals, which Bates (1989) referred to as a "journal run." Hand-searching is a process used to identify relevant articles by reading volumes of a highly relevant journals cover to cover. A study that compared electronic searching and hand-searching revealed that hand-searching successfully identified 92 to 100% of relevant studies for a review (Hopewell et al., 2002). The domain experts in the medical group partitioned the hand-searching tasks between them to ensure that all journals they considered important to the field were reviewed. This finding leverages the tacit knowledge of a discipline that Ellis (1989) observed during his study of experts.

As the collection of articles increased, the medical group started to use the fourth search strategy, which was to propose new articles from the references within existing articles. After the medical librarian retrieved the full text, she applied the group-defined inclusion criterion before circulating a printed copy of the new article to the group members. The medical librarian maintained a database of all articles considered throughout the study. In addition to facts extracted from an article, each group member provided the librarian with additional references from the article that they considered relevant. The librarian would retrieve the full text of each article, and if it satisfied the inclusion criterion, she would add it to the corpus. Although some of their chosen databases would have enabled this user group to search the citations in an automated way (e.g., using Web of Science), the medical group used only a manual technique to search the references within their current collection.

The fifth and last search strategy used by the medical group was to ask for additional references from domain experts. This step is analogous to the author search identified by Bates (1989). The domain experts in the medical group drew on their expertise to identify other domain experts, then provided each new domain expert with a bibliography of the near-complete collection and a request for additional references. This fifth step is critical to the retrieval task because in addition to obtaining a comprehensive collection, asking additional domain experts for references enables the group to identify "gray literature."[3] Gray literature is a collection of unpublished studies that explores the subject of the systematic review. Obtaining gray literature is critical to the systematic review process because published articles are systematically different from articles that are not published. For example, published articles typically show larger effect sizes than do unpublished studies (McAuley, Pham, Tugwell, & Moher, 2000). If epidemiologists did not consider unpublished studies, then higher estimates in each of the published studies would lead them to overestimate risk. Despite its importance, gray literature is difficult to obtain and is thus included in only 33% of systematic reviews (McAuley et al., 2000). Researchers have established trial repositories to ease the identification of gray literature, such as TrialBank (Sim, Owens, Lavori, & Rennels, 2000) and the Cochrane Central Register of Controlled Trials (Dickersin et al., 2002).

In addition to the five methods used by the medical group to identify relevant articles, researchers in medicine have shown that sophisticated queries can improve retrieval performance. Hopewell et al. (2002) found that a simple MEDLINE search identified only 55% of the relevant studies for a review. In contrast, the Cochrane Collaboration's Highly Sensitive Search Strategy increased the successful retrieval rate (i.e., recall) to 80% (Robinson & Dickersin, 2002).

Extraction. The second critical task conducted by both the medical and public health groups was extraction. In contrast to Ellis's (1989) description of extraction that involves identifying an article from a periodical, conference proceeding, and so on, we use the definition of extraction developed by the National Institute of Standards and Technology (NIST) as part of the TIPSTER Text program. The NIST (2004) definition of extraction is "The selection of specific types of information from text, e.g. person name, place names, companies, organizations, or relationships between text entities." Thus, our extraction task couples tightly with user-provided context information.

Unfortunately, we were unable to observe directly either the medical or the public health group as they extracted. The

[3]Gray literature also is call "bottom-drawer" literature (Alderson et al., 2004).

DOI: 10.1002/asi

medical group extracted information outside of our regular meetings, and the public health group had completed the extraction task for both of their reviews prior to our study; however, the director of the public health group estimated that she spent 3 hr per article (T. Tengs, personal communication, January 2002) to extract the required context information required for the study on smoking and impotence.

Both the medical and public health groups distinguished between information that is explicit in an article and information that they could derive from an article. The public health group referred to explicit information in an article as facts that "appeared in black and white." This distinction is important because explicit facts are more reliable than are derived facts. The information extracted comprised both facts that appeared in black and white and the derived facts.

Verification. The accuracy of information extracted from each article is critical to the reliability of the systematic review findings. Thus, it is not surprising that our observations and interviews revealed the third task, verification. The medical group used the following process to verify the manually extracted information items. Three group members independently extracted data from each article. Once extracted, the three group members met to establish agreement regarding the response. If the three members could not resolve a disagreement, a fourth group member considered the justifications for each response and made the final decision. The public health group used a different methodology to verify the extracted facts. One group member would extract the information, and another group member would verify the proposed fact by comparing the fact with the original document. The public health group also verified facts by having two group members check a third group member's proposed extracted fact.

Despite our user populations' shared understanding and formal qualifications, each group verified the information extracted. This finding is consistent with the authorities on systematic review methodologies (Higgins & Green, 2005; Sutton et al., 1998). Hripcsak and Wilcox (2002) reported that ". . . aggregated responses, while not perfect, are more reliable than any single expert's response and may serve as a reasonable reference with which the systems may be compared" (p. 2).

Analysis. The fourth task that emerged from our data was analysis. During the analysis task, scientists apply qualitative or quantitative methods to the verified extracted facts to identify and resolve conflicting evidence within the collection of studies. Although one might consider the analysis task the only task required to integrate findings, the analysis chapter of the Cochrane Handbook (Higgins & Green, 2005) states empathically "Do not start here! Please consult Sections 2 to 6 before reading this Section" (pp. 97).

In the systematic review literature, both a qualitative and quantitative analysis require that a quality score be assigned to each article. The quality score captures both the presence or absence of information and a weight that indicates the importance of each fact. For example, the information item "*Examiners blinded to clinical presentations*" had a quality score of 8 while the information item "age and ethnicity of patients" had a quality score of 1. The quality score of an article is the total weighted score of all information items. Once the group had assigned a quality score to each article, they grouped the articles based on quality and discussed findings with respect to each quality cluster. Scientists account for redundancies and contradictions by emphasizing results reported in high-quality studies.

In contrast to the medical group, the public health group's analysis was quantitative. They combined evidence pertaining to the relationship between smoking and impotence using a randomized-effect meta-analysis (see Background section). The result of their analysis is both a visual summary comprising a box plot and a quantitative summary of the effect size, with a unitless metric that captures the strength of associations.

Process-Level Information Behaviors

Our data revealed two information behaviors that spanned their entire synthesis activities: iteration and collaboration. These behaviors have important implications to the development of systems that would support collaborative information synthesis. In this section, we describe the recurring themes surrounding the iterative and collaborative behaviors that we observed.

Iteration. Iteration is well reported in studies of information-seeking behavior. For example, Dervin (1983) stated that "effective 'circling of reality' is not only desirable (i.e., valued) but necessary given the considerable body of evidence showing what happens to systems unable to assess and respond flexibly to changing reality" (p. 7). Although we expected iteration, we anticipated that domain expertise and familiarity with the systematic review process would minimize iteration by this user population. In contrast, our data revealed that both the medical and public health groups iterated within each task and between the tasks. In this section, we outline the nature and extent of the iteration that occurred both within and between the critical tasks.

The retrieval-task section provides an example of how the medical group iterated. Specifically, the medical group repeatedly employed different retrieval strategies until they were sure that their collection of evidence was comprehensive. The public health group also used a multi-strategy approach to retrieve articles for both their meta-analysis and database maintenance projects. Within-task iteration also occurred during the development of the extraction worksheet, a component of the extraction task. During the pilot phase, scientists worked together to articulate the information required on the extraction checklist. They then extracted information from a set of articles and refined the checklist worksheet. Throughout the analysis task, scientists applied several combinations of data transformations and analysis methods before arriving at a combination that best explained the evidence reported in the literature.

Iteration also occurred between tasks. For example, the medical group iterated between the retrieval and analysis tasks. Their initial goal was to use quantitative techniques during the analysis task; however, they were unable to retrieve a sufficient number of randomized clinical trials to employ a quantitative approach. Thus, the medical group changed their analysis to a qualitative method. In addition to the kind of analysis, the medical group's hypothesis projection evolved as they uncovered more information about the number and kind of studies that were available (see Table 2). The public health group also iterated between the retrieval and analysis tasks. Their preliminary goal had been to understand the relationship between impotence and the joint probability of smoking and alcohol consumption; however, an insufficient number of articles reported the joint probability of smoking and alcohol consumption. Thus, the public health group limited the scope of their study to tobacco consumption. In each of these cases, the scientists required retrieval-task completion before they could predict either their analysis method or research question accurately. This finding suggests that providing a summary of the information in each of the available studies would enable the scientists to establish the nature and scope of their analysis more easily than would the current manual methods.

In addition to their iteration between the retrieval and analysis tasks, the public health group iterated between the retrieval and extraction task to ensure that each article reported a population that was located in the United States, and provided either the number or rate of impotent men who consumed tobacco. This finding suggests that accurate extraction technologies and a tighter integration between the retrieval and extraction tasks would reduce the number of articles requiring review and thus accelerate the synthesis activity. This finding is consistent with a study on information behaviors by McKeown et al. (2001) which showed that users do not distinguish between the retrieval and extraction tasks.

Collaboration. In addition to iteration, our data revealed that the synthesis is not an individual but rather a collaborative activity. During the retrieval task, all members of the medical group added, removed, and modified search terms to ensure that their article collection was comprehensive. Although the domain experts in the medical group were the most active participants in the search-term discussion, other group members contributed by tailoring terms to each bibliographic database and by forcing the domain experts to provide definitions for terms that were commonly used in the spinal-manipulation literature. During the extraction task, the medical group members discussed the assignment of weights to each extracted fact. The group continued to refine the hypothesis projection and to identify additional context information until they reached consensus.

The majority of the medical group's consensus building took place in person during methodology meetings, in which all group members were present. Although our data did reveal that the group used e-mail to disseminate meeting minutes, it was rare that they used e-mail to reach a consensus. In contrast to the intergroup discussions, the intragroup discussions did not take place in person but were initiated with external domain experts using e-mail. The external domain expert then reviewed the citation list, extraction worksheet, and manuscript, and provided feedback using e-mail.

In contrast with the medical group, collaboration was not as prominent in the public health group. We attribute this difference to the differing levels of expertise between the members of each group. With the exception of one member who held a master's degree, all members of the medical group held doctoral-level degrees. The qualifications of the public health group, however, ranged between a high-school diploma and a doctoral degree. In the medical group, each member contributed expertise to different activities during their synthesis activity. For example, the domain experts in the medical group played a major role during the formulation of the hypothesis projection to ensure that it did not duplicate a previous study. The group members, who were experts in complementary and alternative medicine but not experts in the topic of analysis, provided a "third-party" perspective; thus, they ensured that the hypothesis projection was clear to nonexpert readers. The medical librarian provided searching expertise and maintained a database of articles. The statistician played an active role when defining statistical context information and assigning appropriate weights to the extraction worksheet. In contrast to the medical group, where each member played an active role during the project, in the public health group the need for training influenced the role that each group member could play. Specifically students were learning about the process rather than actively participating.

To see if collaboration was idiosyncratic to the two groups that we observed, we identified meta-analyses in MEDLINE that were published between 1990 and 1996 in the following journals: *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] Of the 147 studies that satisfied these criteria, 132 had multiple authors. Thus, a large portion of multi-authored papers supports our finding that the systematic review process is a collaborative activity.

Conclusion

In this article, we have introduced the CIS model that captures the salient information behaviors of the academic scientists that we observed. Although the synthesis activities that we observed capture just one of the day-to-day activities that consumes a scientist's time, however, the effort required to conduct this activity as well as the important role that current, accurate synthesized evidence plays in identifying new

[4]Journal choice based on a previous empirical analysis of meta-analysis (Engels, Schmid, Terrin, Olkin, & Lau, 2000).

DOI: 10.1002/asi

discoveries suggests that the study of synthesis activities demands our immediate attention. Furthermore, we must consider the limited scalability of existing manual techniques in the contexts of soaring quantities of new information.

The agricultural industry developed the meta-analytic techniques used by the biomedicine community that we observed. A variety of scientific disciplines, such as chemistry, psychology, education, and business, use meta-analytic techniques, which suggest that these populations may exhibit similar information behaviors to those that we observed. By observing an expert rather than a novice user population, we are confident that the CIS model reflects ideal synthesis behaviors and anticipate that nonscientific user populations who operate in an information-intensive environment will find the our model valuable.

We observed that during synthesis activities, scientists provide two information constructs—the hypothesis projection and context information—and that they engage in four tasks—retrieval, extraction, verification, and analysis. We observed that synthesis is a collaborative rather than an individual activity and that a scientist will iterate both within and between the critical tasks based on new information.

As the quantity of information at our fingertips continues to exceed human-processing capacity, the importance of information systems that integrate findings will continue to increase. In contrast to an optimal retrieval system that provides a user with all relevant documents, an optimal synthesis system provides a user with an accurate overall finding that reflects contradictory and redundant evidence that is inevitable when multiple studies report the same phenomena. The CIS model introduced in this article provides insight on one method used by experts to integrate vast quantities of evidence and resolve differences. Understanding this process is just the first step toward developing information systems that support synthesis activities. In Part 2 of this study (Blake & Pratt, in press), we provide additional details about the user population, compare and contrast existing cognitive and information models that have inadvertently detected synthesis activities, and provide five specific recommendations for the development of CIS systems.

References

Allen, E., & Olkin, I. (1999). Estimating time to conduct a meta-analysis from number of citations retrieved. Journal of the American Medical Association, 282(7), 634–635.

Asilomar Working Group. (1996). Checklist of information for inclusion in reports of clinical trials. Annals of Internal Medicine, 124(8), 741–743.

Bates, M.J. (1976). Rigorous systematic bibliography. RQ, 16, 7–26.

Bates, M.J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review, 13, 400–412.

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., et al. (1996). Improving the quality of reports on randomized controlled trials. Recommendations of the CONSORT study group. Journal of the American Medical Association, 276(8), 637–639.

Blake, C., & Pratt, W. (in press). Collaborative information synthesis II: Recommendations for information systems to support synthesis activities. Journal of the American Society for Information Science and Technology.

Clarke, M. (2002). The Cochrane Collaboration: Providing and obtaining the best evidence about the effects of health care. Evaluation & the Health Professions, 25(1), 8–11.

Davies, H.T.O., & Crombie, I.K. (1998). What is a meta-analysis? Hayward Medical Communications, 1(8), 1–8.

Dervin, B. (1983, May). An overview of sense-making research: Concepts, methods and results. Paper presented at the annual meeting of the International Communication Association, Dallas, TX. Retrieved from http://communication.sbs.ohio-state.edu/sensemaking/art/artdervin83.html

Dervin, B., & Nilan, M. (1986). Information needs and uses. In M.E. Williams (Ed.), Annual Review of Information Science and Technology (ARIST) (Vol. 21, pp. 3–33). Amsterdam: Elsevier.

Dickersin, K., Manheimer, E., Wieland, S., Robinson, K.A., Lefebvre, C., & McDonald, S. (2002). Development of the Cochrane Collaboration's central register of controlled clinical trials. Evaluation & the Health Professions, 25(1), 38–64.

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

Ellis, D. (1993). Modeling the information-seeking patterns of academic researchers: A grounded theory approach. Library Quarterly, 63(4), 469–486.

Ellis, D., & Haugan, M. (1997). Modeling the information seeking patterns of engineers and research scientists in an industrial environment. Journal of Documentation, 53(4), 384–403.

Elstein, A.S., Shulman, L.S., & Sprafka, S.A. (1978). Medical problem solving: An analysis of clinical reasoning. Cambridge, MA: Harvard University Press.

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

Funk, M.E., & Reid, C.A. (1983). Indexing consistency in MEDLINE. Bulletin of the Medical Library Association, 71(2), 176–183.

Higgins, J.P.T., & Green, S. (Eds.). (2005). Cochrane handbook for systematic reviews of interventions 4.2.5 (updated May 2005). In The Cochrane Library, Issue 3, 2005. Chichester, UK: John Wiley & Sons.

Hopewell, S., Clarke, M., Lefebvre, C., & Scherer, R. (2002). Handsearching versus electronic searching to identify reports of randomized trials. The Cochrane Database of Methodology Reviews 2002, Issue 4.

Hripcsak, G., & Wilcox, A. (2002). Reference standards, judges, and comparison subjects: Roles for experts in evaluating system performance. Journal of the American Medical Informatics Association, 9(1), 1–15.

Lipsey, M.W., & Wilson, D.B. (2000). Practical meta-analysis (Vol. 49). Newbury Park, CA: Sage.

McAuley, L., Pham, B., Tugwell, P., & Moher, D. (2000). Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? Lancet, 356(9237), 1228.

McKeown, K., Chang, S.F., Cimino, J., Feiner, S., Friedman, C., Gravano, L., et al. (2001, June). PERSIVAL, a system for personalized search and summarization over multimedia healthcare information. Paper presented at the 1st Joint Conference on Digital Libraries, Roanoke, VA.

Merriam-Webster's Collegiate Dictionary. (11th ed.). (2004). Springfield, MA: author.

Moher, D., Schulz, K.F., & Altman, D.G. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. Journal of the American Medical Association, 285(15), 1987–1991.

Murphy, L.S., Reinsch, S., Najm, W.I., Dickerson, V.M., Seffinger, M.A., Adams, A., et al. (2003a). Searching biomedical databases on complementary medicine: The use of controlled vocabulary among authors, indexers and investigators. BMC Complement Alternative Medicine, 3(1).

Murphy, L.S., Reinsch, S., Najm, W.I., Dickerson, V.M., Seffinger, M.A., Adams, A., et al. (2003b). Spinal palpation: The challenges of information retrieval using available databases. Journal of Manipulative and Physiological Therapeutics, 26(6), 374–382.

National Institute of Standards and Technology. (2004). TIPSTER text program glossary of terms. Retrieved July 26, 2004, from http://www.itl.nist.gov/iaui/894.02/related_projects/tipster/gloss.htm#sectE

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

DOI: 10.1002/asi

Rescher, N. (1978). Peirce's philosophy of science critical studies in his theory of induction and scientific method. Notre Dame, IN: University of Notre Dame Press.

Robinson, K.A., & Dickersin, K. (2002). Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. International Journal of Epidemiology, 31(1), 150–153.

Sim, I., Owens, D.K., Lavori, P.W., & Rennels, G.D. (2000). Electronic trial banks: A complementary method for reporting randomized trials. Medical Decision Making, 20(4), 440–450.

Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.D., Rennie, D., et al. (2000). Meta-analyses of observational studies in epidemiology. Journal of the American Medical Association, 283(15), 2008–2012.

Sutton, A.J., Abrams, K.R., Jones, D.R., Sheldon, T.A., & Song, F. (Eds.). (1998). Systematic reviews of trials and other studies (Vol. 2). York, United Kingdom: York Publishing Services.

Tengs, T., & Osgood, N.D. (2001). The link between smoking and impotence: Two decades of evidence. Preventive Medicine, 32(6), 447–452.

DOI: 10.1002/asi