The Basis for Using the Internet to Support the Information Needs of Primary Care

Edward E. Westberg, MS, Randolph A. Miller, MD

Abstract Synthesizing the state of the art from the published literature, this review assesses the basis for employing the Internet to support the information needs of primary care. The authors survey what has been published about the information needs of clinical practice, including primary care, and discuss currently available information resources potentially relevant to primary care. Potential methods of linking information needs with appropriate information resources are described in the context of previous classifications of clinical information needs. Also described is the role that existing terminology mapping systems, such as the National Library of Medicine’s Unified Medical Language System, may play in representing and linking information needs to answers.


Over the last two decades, studies have enumerated the information needs of physicians engaged in clinical practice, including primary care. The emergence of the Internet and widespread adoption of the World Wide Web (the Web) have improved clinician’s access to information resources. The Internet and the Web provide clinicians with both a ubiquitous, standardized system interface and a variety of Web-based materials. This review focuses on the potential of Internet-based resources to address the information needs arising from primary care.

The Nature of Primary Care

Primary care is the “provision of integrated, accessible, health care services by clinicians who are accountable for addressing a large majority of personal health care needs, developing a sustained partnership with patients, and practicing in the context of the family and the community,” Noble et al. emphasize the broad spectrum or “constellation” of health care services comprised by primary care practice. Examples of these services include hospital services, emergency medical services, public health, counseling, and home care services.

The art of clinical practice concerns the overall management of a patient’s well-being. As described by Levinson, this role includes the task of information management:

The physician is an information manager who acquires, processes, stores, retrieves, and applies information related to 1) individual patient history and clinical course, 2) diagnostic and therapeutic protocols, 3) disease patterns in patient populations, 4) functioning of the health care system, and 5) the vast store of published knowledge. Little occurs in the clinical encounter that is not in some way related to obtaining, processing, or applying information. Optimal performance of clinical informational tasks has for years exceeded the cognitive capability of the human mind.

Observational Studies Describing the Information Needs of Clinical Practice

When faced with clinical decisions, primary caregivers (and other clinicians) must recognize when it is important to seek additional information rather than
Gorman’s review of information needs concludes that “questions about optimal patient care are frequent, with many questions occurring each day for a typical physician.” As part of their analysis of information needs expressed by general internists, Osheroff, Forsythe, and colleagues categorized questions’ status as being currently satisfied (clinician recognizes question and knows answer), consciously recognized but unsatisfied (question recognized, answer unknown), and unrecognized (clinician should perceive information need exists but does not, and does not know answer). Published estimates of the incidence of information needs during clinical practice range from 1 question generated for every 15 patients in primary care settings to roughly 1.4 questions per patient on a given day in inpatient settings. It is not clear whether the observations from primary care settings covered “unrecognized” information needs.

Osheroff, Forsythe, and colleagues introduced a two-by-two matrix that characterizes clinical knowledge along two axes: formal to informal, and general to specific. Examples of formal knowledge include the peer-reviewed literature and databases from controlled scientific studies. Unwritten, common practices followed at a clinical site and population-based information from local, noncontrolled clinical data repositories represent examples of informal knowledge. General knowledge, whether formal (textbooks, literature) or informal (untested guidelines), is available widely and applies to categories of patients. Examples of local knowledge include the clinical findings in the chart of a patient and guidelines developed at a single clinic. Gorman classified the types of information used by clinicians into five categories: patient data, population statistics, biomedical knowledge, logistic information, and social influences. Thus, from a clinician’s perspective, lack of desired information of any of these types may constitute an information need.

An unmet information need is an information need that has not been answered when the clinician makes a decision about the patient. Because of the ever-increasing size of the biomedical literature and the complexity of modern health care practices, clinicians could spend hours to weeks reading texts and seeking expert opinions for each patient they encounter. During a busy primary care practice, clinicians must constantly trade off providing care to more patients versus addressing information needs arising from patient encounters. Time pressures of clinical practice make it difficult to answer information needs as they arise and may explain why, after completing residency training, a physician’s knowledge of medicine tends to decline over time.

Other barriers to addressing information needs include limited access to resources, the cost of resources, difficulty learning or using many resources, poor organization of resources, and variable quality of information. Quantitative and qualitative analyses of unmet information needs have been performed in settings ranging from large academic institutions to small primary care offices and clinics. Unmet information needs range from 0.12 to 5.2 unanswered questions per half-day, depending on the practice setting and sampling methodology.

The influence of unmet information needs on patient outcomes is unknown. Even though every unmet information need potentially compromises patient care, a good clinician is expected to determine whether meeting a given information need is mandatory or critical at the time of a specific patient encounter. However, Williamson et al. surveyed primary care practitioners in the United States and found that they require “substantial help” meeting their information needs. This finding is significant, because it showed that “... physicians face a serious problem in their effort to keep current with recent medical advances.” Bankowitz et al. demonstrated that diagnostic uncertainty influences resource utilization during patient evaluation.

In summary, information needs are numerous and exist in many forms, in primary care as well as other clinical settings. Studies of information needs over the last two decades underscore the persistence of the problem and imply that for whatever reasons, evolving information resources have not yet reduced primary caregivers’ unmet information needs significantly.

**Sources of Information Currently Available to Primary Care**

Observational studies and self-reflective surveys suggest that the main resources that clinicians, including primary care providers, use to satisfy information needs include colleagues, tertiary literature (such as...
textbooks), primary literature (such as original research in clinical journals), and continuing medical education. Less widely accessible, traditional sources of information also include health science library-based and bibliographic resources, academic health care centers, and clinical consultations and referrals. Less well established but increasingly useful sources of information include clinical computer software applications, telemedicine applications, and the Web.

**Health Science Libraries and Bibliographic Resources**

A significant number of primary care settings are distant from facilities that provide health information. Many primary care practices are in underserved areas, are understaffed, and generally do not employ in-house subspecialists for consultation services.

**Health Science Libraries**

Health science libraries offer a variety of services that can help primary care providers address information needs, including traditional print-based textbooks and journals, interlibrary loans, librarian-mediated literature searches, computer training on bibliographic searching software and clinical software tools, alerting and document delivery services, easy access to bibliographic citation databases (e.g., MEDLINE), and local access to electronic “full-text” journals and textbooks.

Health science libraries often have sufficient resources to mount a representative sampling of clinical textbooks distributed in electronic (usually CD-ROM) format; few other institutions do so. Libraries or large networked office practices can make such offerings accessible via an intranet. Products utilizing electronically based reference sources provide superior searchable interfaces compared with print-based versions of the same sources. For example, the Harrison’s Plus product contains electronic versions of Harrison’s Principles of Internal Medicine and the U.S. Pharmacopoeia’s Drug Information for the Health Care Professional. Examples of other electronic textbooks include Scientific American Medicine, Physician’s Desk Reference (PDR), and Nelson’s Textbook of Pediatrics, among many others.

The value of all of these library-based services to clinical practice, including primary care, has been studied in academic health care centers, urban hospitals, urban community health centers, and rural areas. Although clinicians vary by practice setting in the resources they access, there is positive support for utilization of these library-based services.

**Clinical Medical Librarian Programs**

Over several decades, health science librarians at selected sites have served successfully as information consultants for patient care. Initiated by Gertrude Lamb, clinical medical librarian (CML) programs began in the early 1970s with the goal of directly increasing the relevance of the health care literature to care providers. In the traditional CML model, academic medical librarians attend either clinicians’ morning report (when new cases admitted to the hospital are discussed) or house staff teaching rounds on a regular basis. After interacting with the health care team, they determine information needs and place relevant literature references on the patients’ charts (a process called LATCH, for “Literature attached to charts”). This allows the librarian to provide specific, case-related information to support patient care. In a more recent model for CML, librarians participate as members of ward teams rather than simply attaching literature to charts. Overall, CML programs are well accepted among care providers but they have been criticized for being labor-intensive and expensive.

In recent years, the proliferation of end-user–friendly information resources (such as the NLM’s Grateful Med) caused a shift in biomedical librarians’ role from information mediators to knowledge workers. This “revolution” encourages librarians to work in settings beyond the library’s walls and to seek new, proactive roles for development and delivery of information products and services to the clinical practice, biomedical research, and patient communities. For example, one service extends the traditional quality filtering of the literature provided by librarian-mediated searches. Instead, in response to an information request, clinically knowledgeable librarians read, filter, and synthesize key articles and produce a concise written summary relevant to a specific clinical case. By entering the results of their consultation into a growing knowledge base, librarians can better support related inquiries in the future.

Reusable knowledge resources created by such efforts can help support the information needs of primary care practitioners and patients themselves.

**Bibliographic and Information Retrieval Software**

The National Library of Medicine has developed MEDLINE-based resources such as PubMed, Grateful Med, and LOANSOME DOC that help primary caregiv-
ers benefit from biomedical library journal contents from the clinicians’ practice settings. In addition, computers equipped with a CD-ROM drive or a modem can access bibliographic search software, such as Ovid, Aries Knowledge Finder, SilverPlatter, and PaperChase.

The bibliographic citation and full-text information retrieval programs developed by developers and vendors vary with respect to domain (e.g., infectious diseases or pharmacology), content (e.g., citations or full-text articles), user interface (e.g., free-text input or codified query terms), and indexing and retrieval methods (e.g., statistical or concept-based). As a result, the quality of citations retrieved by various MEDLINE-based search interfaces varies considerably.

With respect to health care providers, studies indicate that MEDLINE and its derivatives can be helpful for answering clinical questions. However, finding specific answers to questions can be time-consuming and expensive, in part because of the effort required to sift through a sometimes large set of relevant publications. In that regard, some have likened MEDLINE searching to attempting to drink water from a fire hose.

**Academic Health Centers**

Academic health care centers maintain a large array of resources that help clinicians keep current with biomedical advances and more effectively care for their patients. As noted earlier, primary caregivers lack ready access to current journals, textbooks, and multimedia resources like those held by the libraries of academic health care centers. Also, among the commonly sought-after sources of new information for primary caregivers are the continuing medical education courses offered by academic health centers.

**Academic Consultation and Referral**

Primary care providers practicing outside large health care centers often lack easy access to subspecialists. They often seek their assistance in the form of referrals and consultations, although communication between referring physicians and consultants can also be problematic. Consultation and referral patterns vary considerably, as documented by studies in the United States and the United Kingdom. Factors linked to variation include availability of qualified consultants, diagnostic certainty, referring-provider training, and specialty, and reimbursement plan. It is not unusual for a referring physician to receive little or no relevant feedback related to the information request that prompted a referral.

Consultation and referrals from primary caregivers to clinical specialists do not uniformly involve academic medical centers. For the purposes of this review, the authors have listed consultation and referral under academic health centers because such centers generally provide the widest range of consultative services in a region. The collective academic knowledge and pragmatic skills of the faculty and staff at academic health centers are their most valuable resource. Health centers are experimenting with a number of new approaches to replace the traditional, labor-intensive, one-on-one phone call from a primary caregiver to an academic colleague. For example, the Medical Information Services via TelePhone (MIST) network at the University of Alabama in Birmingham (UAB) provides thousands of rural physicians and other health care professionals with toll-free 24 × 7 access to free consultation services with UAB Medical Center faculty and staff. The success of MIST has motivated other institutions to develop similar programs.

**Academic Health Center Telemedicine Initiatives**

Telemedicine is “the use of electronic information and communication technologies to provide and support health care when distance separates the participants.” Because they generally have sufficient resources as well as extended referral networks, academic health centers have initiated or coordinated the majority of telemedicine initiatives in civilian settings. As opposed to Web-based “generic” informational resources (discussed later in this article), telemedicine consultations are patient-specific and patient-focused.

The definition of telemedicine encompasses a wide variety of clinical applications, including telephone-based systems for voice support, relatively low bandwidth systems for sharing textual and multimedia data, and higher bandwidth systems transmitting interactive video. While some telemedicine systems support clinical consultations across a variety of specialties, other telemedicine applications specialize in individual domains, such as teleradiology, telepsychiatry, and teleradiology. The use and growth of teleradiology systems have led to reimbursement by Medicare.

A number of telemedicine systems at least partially address the information needs of primary care. Georgia’s GaN, and West Virginia’s CONSULT are two older, established statewide networks whose participating “spoke” sites receive MEDLINE access, e-mail, and consultative services. Both networks give
community physicians the opportunity to channel information requests to trained "information gatekeepers" at remote sites, such as regional hospitals. The more recent, four-state regional IAIMS effort developed by the University of Washington implements telemedicine for education and consultation.\textsuperscript{175,176}

A new, rapidly growing form of telemedicine extends support directly to patients, as an adjunct to services provided by primary caregivers. For example, Brennan et al.\textsuperscript{142–145,177–180} developed ComputerLink, an electronic network to assist family members (and others) who care for patients with Alzheimer’s Disease. Participants used home computer terminals with modems to access ComputerLink. The service offered three main functions: electronic encyclopedia, e-mail, and a decision support system.\textsuperscript{180} ComputerLink enhanced caregivers’ decision-making confidence\textsuperscript{142} and was useful in discovering the types of support required by patients and caregivers.\textsuperscript{142}

At present, a number of unresolved issues hinder use of telemedicine in primary care settings: licensing across state lines, lack of standards, difficulty of training users and maintaining equipment, reimbursement policies, issues of patient confidentiality, and the cost of telecommunication infrastructure.\textsuperscript{181–183} The time required to adequately train users is substantial.\textsuperscript{177,185} Many telemedicine efforts initiated as demonstration projects assessed the technical feasibility of electronic communications to support patient care,\textsuperscript{181} but not its cost-effectiveness.\textsuperscript{184–185} Telemedicine systems based on real-time, state-of-the-art videoconferencing equipment are costly and may not be practical for most institutions.\textsuperscript{181,182} Evidence suggests that simple, less sophisticated store-and-forward methodologies may realize cost savings without compromising quality by relaying stored patient data, sound, and images to remote sites for later "asynchronous" consultation.\textsuperscript{138}

Ultimately, the federal High Performance Computing and Communications program and the Next Generation Internet\textsuperscript{193,194} and Internet\textsuperscript{2}\textsuperscript{195} initiatives will provide substantial support for use of the national information infrastructure by the health care community, including implementation of sophisticated telemedicine applications.\textsuperscript{196–199}

Collectively, the print, electronic, and human information resources at academic health care centers can be useful to primary care. However, barriers to effective and widespread use of these resources—related to ease of access—remain.\textsuperscript{43} It is difficult for the remote practitioner to know who or what is available and how to best make use of services from a distance.\textsuperscript{43,58}

Clinical Computer Software Applications

A variety of clinical computer software products have been developed to support clinical decision making. For the purposes of primary care, these applications can be broadly classified into three categories: bibliographic and full-text information retrieval systems (discussed earlier)\textsuperscript{200,201} clinical decision-support systems (CDSS),\textsuperscript{13,202} and clinical information systems (electronic medical record systems, or EMRSs). A thorough discussion of CDSS and EMRS is beyond the scope of this review,\textsuperscript{92,203–206} however, a brief overview of their relevance to primary care follows below.

Clinical decision support systems have the potential to provide primary caregivers with useful information regarding diagnosis, therapy, and prognosis. Given a set of patient findings, diagnostic CDSS can, for example, compute and explain differential diagnoses, show relevant laboratory tests sorted by cost, suggest possible workup protocols, and provide links to relevant biomedical literature.\textsuperscript{207,208} Examples of broadband diagnostic CDSS for general internal medicine include Dxplain,\textsuperscript{209} ILIAD,\textsuperscript{210–212} Meditel,\textsuperscript{213} and Quick Medical Reference (QMR).\textsuperscript{214} Rigorous evaluations of CDSS are difficult to conduct.\textsuperscript{13,215–218} Evaluations of Dxplain,\textsuperscript{215,219,220} ILIAD,\textsuperscript{210,215,219,221} Meditel,\textsuperscript{215,219} and QMR\textsuperscript{216,219,222} indicate that no broadly based diagnostic systems perform superiorly in terms of accuracy, but many are able to suggest additional diagnoses not originally considered by users. Providing adequate training to users of such systems is an ongoing, unsolved problem.\textsuperscript{30,221} Overall, CDSSs, when properly used with an understanding of their strengths and limitations, can potentially offer useful advice to primary caregivers presented with complex clinical problems for which they might otherwise seek consultation.

Clinical information systems (EMRSs) promote more effective patient care by reliably and efficiently storing and retrieving patient data—ranging from clinician’s orders to clinical textual reports (e.g., history and physical examination notes, progress notes, nursing notes, discharge summaries, radiology reports, and pathology reports), numerical laboratory results, pharmacy information, billing information, census data, and outcome data.

A number of early studies attempted to integrate microcomputer-based bibliographic citation databases, drug information databases, electronic textbooks, and decision support systems to answer clinical questions for primary care.\textsuperscript{30,35,223,224} One study compared two groups of private practice physicians, nurses, and university-based pharmacists using either BRS Colleague
or Dialog Medical Connection resources for acquiring
drug information.223 Users obtained less than com-
plete information for 70 to 86 percent of their ques-
tions. Study authors found pertinent information for
59 percent of failed user searches. Improper use of
search terms and failure to select all relevant data-
bases constituted the two most significant searching
errors. A separate descriptive study of eight physi-
cians using six commercially available clinical soft-
ware products during a two-week period20 found that
40 percent of the questions arising in daily practice
were fully answered; 32 percent were partially an-
swered; and no useful information was obtained for
28 percent of the questions. The study’s authors de-
scribed inappropriate resource selection by the phy-
sicians and were subsequently able to locate relevant
information for failed searches. In a ten-month study
by Hersh and Hickam,224 a computer workstation was
provided for routine use in a university-based general
medicine clinic. The workstation contained bibli-
ographic search software, full-text textbook searching
software, and decision support software. The authors
concluded that novice searchers could retrieve large
quantities of relevant information when provided
with user-friendly software.

In summary, while microcomputer-based software ap-
lications contain large amounts of useful informa-
tion, significant barriers to the effective retrieval and
application of that information remain in primary
care. Users have difficulty finding the most relevant
resources, are unable to master multiple applications,
and require time-consuming, out-of-the-office train-
ing. Hence, the utility of existing resources is limited.

It has been stated for decades that clinical software
systems will achieve their greatest value when larger
EMRs, CDSs, and other clinical applications are in-
tegrated seamlessly across systems and across sites.
Even in academic health centers, this lofty objective
has been achieved with only partial success (for ex-
ample, alerting225 and critiquing226 during physician
order entry207,228). Access to electronic information
stored elsewhere remains as much an impediment (or
possibly more, because of security issues) to primary
care as is access to paper records stored elsewhere. It
remains problematic to meet those information needs
of primary care related to utilization of patient data
stored at another site.229

The World Wide Web

While it is clear that Web-based interfaces will play
essential roles in delivering easy-to-use systems to pri-
mary care providers, the role of the Web, taken as an
information resource per se, in supporting and ad-
dressing primary care information needs is far less
certain. Care providers can find clinically useful infor-
mation on the Web,230–236 but the time requirements
for such searches can be substantial.237–240 Indeed, critical evaluation of the representation of health information on the Web
is warranted. The Web’s rapid growth241–243 and lack
of controls have led to numerous criticisms, including
poor organization, questionable validity, and ques-
tionable reliability.244,245 These shortcomings effectively
render a substantial amount of Web information unsuitable for direct clinical application.246–251

The best and most widely accepted strategy for use
of the Web to support clinical practice involves locating
and using “anchors” of known high quality. A number of U.S. government agencies, such as the Na-
tional Institutes of Health (NIH)252 and the Food and
Drug Administration (FDA),253 provide useful, reliable
Web-based resources relevant to primary care. For
example, two institutes within the NIH—The National
Cancer Institute (NCI)254 and the National Library of
Medicine (NLM)255—provide free, unrestricted access
to a diverse set of clinically useful information re-
sources. These include the NLM’s PubMed,256 Internet
Grateful Med,257 and Health Services Technology As-
sessment Texts (HSTAT)258—which contain clinical
practice guidelines, quick reference material for cli-
nicians, and evidence-based reports from the Agency
for Health Care Policy and Research (AHCPR)— and
the NCI’s PDQ and Cancerlit databases—which con-
tain information about the cause, diagnosis, preven-
tion, and treatment of cancer.259–261 The FDA, in ad-
dition to other services, provides timely information
on new drug products.262 Academic health care cen-
ters and professional specialty organizations are also
good sources for relevant clinical information.263

Many peer-reviewed journals, such as the New En-
l gland Journal of Medicine264 and the Journal of Fam-
ily Practice,265 have quality full-text articles on the Web.
Other Web sources relevant to primary care include
multimedia textbooks such as the Diagnosis of Pulmo-
nary Embolus,266 published by the Virtual Hospi-
tal266–268; drug review articles such as “Cardiovascular
Drug Reviews”269 from the Medical Sciences Bulle-
tin270; diagnosis and treatment information such as
that published in the Merck Manual of Diagnosis and
Therapy271; forums for asynchronous discussion, such
as USENET newsgroups272–274 and listservs; Web sites
of national voluntary health agencies, such as the Amer-
ican Heart Association275; and organizations that index
Web content such as Medicine in the Matrix276 and
CliniWeb.277,278

No systematic comparative studies have assessed ef-
fects of Web-based information resources on the qual-
ity, accessibility, or cost of primary care. However, demonstrated utility of existing resources (e.g., MEDLINE, CDSS, printed textbooks) prior to their availability on the Web suggests that Web-accessible versions of these resources will also be useful.

A Proposed Model for Internet Use

No single genre of information resource, like those discussed in the previous section, can meet all the information needs in primary care. Primary caregivers must be empowered to utilize all relevant resources efficiently, as they are needed.

Since 1995, the authors have been developing a new model for using the Internet to support the information needs of primary care. This effort has been supported by grants from the National Library of Medicine. The model represents a commonsense synthesis of previous efforts and has been independently proposed by others with minor variations, as discussed later. The model focuses on modern academic health centers as the most logical site to integrate and distribute a wide variety of both electronic and human information resources for primary care. According to the model, academic health centers would mount a Web-based, Internet-mediated triage system to facilitate access to their electronic and human information resources. The triage system would support access to academic centers’ large, well-staffed biomedical libraries and their up-to-date faculty expertise covering clinical subspecialties as well as health services and basic science research and their advanced health informatics projects, including telemedicine. The information resources mentioned earlier are well represented in almost all academic health care centers.

The authors’ model involves three layers of triage. The first layer would use a Web-based secure interface and library-based community outreach techniques to train and remotely connect affiliated primary-care end users to academic health science libraries. Primary caregivers would then, as a first pass, have full access to the electronic resources of the health center library, such as bibliographic and health-related databases, full-text journals, and clinical software applications—nearly equivalent to being on site to address primary care information needs. As part of this process, issues of software licensing must be addressed carefully. This layer is not particularly innovative, in that a number of academic health sciences libraries have mounted significant informatics-related outreach efforts over the last two decades.

The second and third layers of the model involve academic health centers mounting, within the Web-based secure interface of layer one, a software application to field and triage individual practitioners’ information requests that caregivers could not answer using first-layer resources. Users could submit queries in limited natural language format (e.g., one sentence only) or using templates. If the query were submitted in natural language, a parsing system would convert it into a structured format, using an underlying set of templates to represent the possible kinds of information requests. Alternatively, direct entry using templates could circumvent the requirement to parse and make sense of natural language queries. The second-layer triage application would be able to suggest relevant information resources (both Web-based content and dedicated application software product content) for a particular question and present the user with the choice of pursuing various resources further to the point of answer retrieval. Queries could involve general or specific and formal or informal knowledge. For example, an application program might help primary caregivers to identify qualified and available academically based subspecialty consultants and help schedule appointments for patients to see them, as well as improve the suboptimal bidirectional information flows that currently exist.

Automated or semiautomated triage of primary caregivers’ requests could make use of scarce human resources more appropriate and relevant. The third layer of triage in the proposed model would become accessible when users indicated that their request for information had not been adequately answered through use of the first two layers. At this stage, the stored information request would be forwarded electronically to a knowledgeable human “gatekeeper” (e.g., librarian or clinician) for manual triage to appropriate faculty specialists (via e-mail) or to other persons with specialized access to information resources (such as a librarian with access to private institutional databases that do not have public interfaces). These specialists would respond to the primary caregiver’s question via e-mail. Periodic manual audits of such sessions could be used to determine whether electronic systems might have been used to obtain similar results and, if so, why they had not been used. Issues of licensure (ability to give advice across state lines) and reimbursement for faculty and staff time would have to be addressed, as in telemedicine.

Previous Implementation of the Model

Implementing the primary care triage model requires the accomplishment of three main objectives, in a computationally tractable manner: representation of
primary care information needs in a detailed, structured classification; representation of the content of electronic biomedical information resources, as well as topics of human expertise, in a detailed, structured classification; and efficient linkage of specific clinical information needs with appropriate information resources. Health care informatics researchers have already performed substantial work toward accomplishing these objectives. Implementation of the proposed model should take advantage of lessons learned from previous efforts.

Relevance of the UMLS Project to the Model

Most investigations of clinical information needs have reported results in aggregate form (as general schemes) rather than as enumerative taxonomies of what clinicians ask.1–6,15,24–48 Systems designed to service information needs in primary care require a greater degree of categorization than has been reported. Critical missing resources are a standardized, validated, clinically useful classification of information needs suitable for providing Internet-based decision support and a corresponding classification of relevant information resources.

Since no detailed classification scheme for the information needs of primary care was created as the product of observational studies, it is useful to ask whether any existing classification schemes, such as clinical terminology systems or mappings developed as part of implementation projects (rather than derived from observational studies), can be used to represent the information needs of primary care.

Overall, the National Library of Medicine’s Unified Medical Language System (UMLS) is potentially the essential resource related to information resources in primary care because, in essence, it combines many individually important lexical resources now in clinical use.

Overview of the National Library of Medicine’s UMLS Project

The UMLS279–282 was developed to serve as an interlingua283,284 for electronic interchange among disparate clinical and biomedical research systems. The UMLS comprises four evolving knowledge sources: the Metathesaurus, Semantic Network, Information Sources Map, and the SPECIALIST lexicon.285–288 The Metathesaurus is a database of biomedical concepts (and related information) accumulated from more than 40 important controlled vocabularies and classifications in actual use in biomedicine.286 It contains names and semantic information about 500,000 biomedical concepts. The Metathesaurus includes MEDLINE co-occurrence data, which quantify the number of times two terms listed in MeSH (the Medical Subject Headings thesaurus, one of the Metathesaurus component vocabularies) occur together in the literature as the main index terms on the same article.289 The Semantic Network lists semantic types that can be assigned to all Metathesaurus concepts and specifies the types of meaningful relationships that can occur between pairs of semantic types.286,289 The Information Sources Map contains information characterizing the scope and content of hundreds of biomedical information resources. The SPECIALIST Lexicon contains syntactic and semantic information about a subset of the Metathesaurus biomedical concepts (at the term and word levels) for use in natural language processing systems.287,288,291–293

Theoretically, a computer program armed with knowledge from the UMLS could be used to first recognize and then explore possible relationships among concepts in a clinical question expressed in natural language. This information could then facilitate automated or semiautomated methods for mapping the user’s question into a structured representation, which in turn could facilitate determination of relevant information resources via the Information Sources Map (ISM).

Early UMLS Project Work Relevant to the Proposed Triage Model

Before and during the first five years of the UMLS project,279–282 a number of institutions participating in it attempted to address the issue of classifying and answering clinicians’ information needs in a manner potentially relevant to primary care.

In 1985, researchers at the NLM reviewed 2,000 literature search request forms submitted from the National Institutes of Health and created a database of 155 representative queries for experimentation in bibliographic retrieval.294 Although not a formal classification of information needs, this resource contains carefully selected questions covering clinical research, basic science research, and health services research. The database also contains question-specific MEDLINE citations found by an expert NLM searcher and corresponding citation relevancy judgments formulated by a subject matter expert. In addition to experiments in bibliographic retrieval, this database has been successfully used to develop natural processing tools for query interpretation.287,295,296

During 1989–91, Osheroff, Forsythe, and colleagues34,297 developed a coding scheme (a hierarchy
with 103 terminal nodes) for describing general medicine information requests at an academic teaching institution. Topics included questions about disease states (e.g., pathophysiology, specific therapies), about therapy in general (e.g., medications or surgery), and about clinical findings (e.g., differential diagnosis of a given finding). In addition to classifying the subject of each information request, the study documented anticipated sources of responses, the generality of information sought, and the nature of responses required. Of the information needs observed, 52 percent of the questions requested a fact that could have been found in a clinical record; 23 percent were potentially answerable by a library (resources such as textbook, a journal, or MEDLINE); and 25 percent required synthesis of patient information and biomedical knowledge.

From 1989 to 1992, a group at Yale University developed two knowledge-based programs designed to help clinicians find relevant literature references, one in psychiatry, called PsychTopix, and the other in hepatology, known as HepaTopix. The programs suggest possible topics of interest based on a scan for key words in a patient’s computer-based record. Selecting a topic generated an automated MEDLINE search using MeSH logic from the program’s knowledge base. For HepaTopix, two pathologists and a hepatologist created a master outline of 35 key topics. Two hundred and twenty-five subtopics were created for five main liver neoplasm topic areas. Subtopics included clinical manifestations, epidemiology, pathogenesis, signs and symptoms, diagnostic tests, and treatment. The outline format enabled further specification as needed. For example, the epidemiology of alcoholic liver disease was further specified by incidence, genetic susceptibility, and sexual susceptibility. Although these topic areas were not derived empirically, they illustrated a useful method of organizing and representing information needs in a given domain. These projects created a mechanism, in the form of hierarchic topic areas, that linked clinical questions with resources that could provide answers. A major drawback was the labor-intensive reliance on domain experts for creating topic areas. This limited application of the approach to other domains.

From 1991 to 1996, Cimino et al. examined common syntactic and semantic patterns in a collection of clinical questions from three sources: the 1985 NLM collection of questions described above, a collection of questions from a cystic fibrosis research database for experimentation in bibliographic retrieval, and reference queries submitted to the Columbia-Presbyterian Medical Center (CPMC) health sciences library. The goal of Cimino et al. was to identify a set of general-purpose questions, called “generic queries,” which can be tailored to user information needs. They hypothesized that use of generic queries in clinical applications could facilitate determination of users’ information needs and simplify selection of potentially relevant information resources. Combining manual review by expert librarians with natural language processing techniques, they derived 37 generic queries that captured the essence of all user queries in their study. The queries typically involved one or two clinical terms and a relation, such as “What causes X?” and “Does X cause Y?” They developed MEDLINE search strategies for each generic query and integrated this knowledge into a clinical application known as the MEDLINE Button. The CPMC clinical information system reviews data in a patient’s computer-based record and, based on the presence of data about a particular disease or therapy, suggests possible questions for which the system may provide information. A particular strength of this effort lies in its generic approach to representing basic information needs and in tailoring the approach to a specific situation.

During 1991–93, a group at Massachusetts General Hospital and Harvard University analyzed physician-generated questions in an ambulatory care setting, in order to identify useful information resources to include in their Interactive Query Workstation (IQW). They determined the types of clinical questions physicians ask, how questions are generally stated, and relevant information resources for answering questions. They collected 69 questions from three physicians, whose information needs were studied by review of the clinical records of 15 of their patients. Physicians also identified the item in each record most closely related to their question. The researchers mapped these “key terms” into the following seven COSTAR categories: physical examination findings, problem lists, medications, nonmedication therapy, laboratory results, procedures, and administrative aspects. All key terms came from three categories: medications, laboratory results, and problem lists. Further analysis quantified the extent to which key terms, query words, and query concepts mapped to the UMLS. The authors concluded that identification of concepts using the (early versions of) UMLS was “not usually sufficient” in describing information needs and that creation of a vocabulary that can represent relationships between concepts was necessary.

Overall, these early explorations and other investigations contributed valuable insights into how
clinical questions may be logically organized and productively integrated into information systems. A common theme is that investigators felt it necessary to develop query representation schemes because suitable representations were not available. They had partial success in linking clinical questions with possibly relevant information resources. These efforts involved building integrated “front ends” for distributed information resources such as bibliographic citation databases, clinical textbooks, and clinical diagnostic decision support systems. Investigators either used home-grown query languages or attempted to adapt existing terminology mapping systems such as the UMLS for such purposes. It is difficult to determine whether their classifications could be extended for providing Internet-based decision support in primary care. Although no common approach yet exists to classify information needs to support computer-assisted query systems, commercial publishers have begun to exploit electronic media, combining a number of Web-based electronic bibliographic and other resources into integrated systems.

Evaluations of the Relevance of UMLS to the Proposed Triage Model

Miller et al., at Yale University, contributed to early applications involving and evaluating the ISM. They reported great difficulty making practical use of the ISM, particularly in two main areas: Encoding the subject content of electronic resources was extremely difficult using current coding schemes; and achieving seamless, cross-platform access to heterogeneous resources was difficult with network communication software. Modifications of the ISM, including use of the Web and application of conceptual graph theory, have been explored to work around these problems.

Because the initial design of the ISM predates the dramatic rise in Internet connections and invention of the Web, its methods for description and access of available machine-readable information resources was insufficient for keeping up with the explosive growth in Web-based information. The dramatic advances in Internet and Web-based technology have had a “temporarily disruptive effect” on UMLS efforts to achieve its long-term goal of defining a method of describing available machine-readable information resources to support automated selection and retrieval from relevant resources.

It is uncertain whether source vocabularies within the UMLS can collectively provide adequate coverage of clinical concepts in primary care queries. In 1987, Masarie and Miller found approximately 50 percent of the words in a medical chart mapped to MeSH. The study by Chute et al. of major clinical coding systems for representing patient information found that UMLS 1.3, ICD-10, SNOMED III, READ V2, ICD-9-CM and CPT—the latter two in prevalent use in the United States as well as constituting a portion of the UMLS source vocabularies—failed to capture “substantial” clinical content. Although these studies document inadequacies of individual and combined vocabularies, a study of the 1997 UMLS by the NLM found that the combination of source vocabularies represents the “majority of the terminology needed to record patient conditions.” It is important to remember that comparative studies of evolving lexical terminologies have a “useful” half-life of at most a few years. In general, each revision of a terminology includes enough term additions, deletions, and other modifications to effectively render previous comparisons among it and other terminologies invalid.

Several additional issues potentially hamper application of the UMLS for representing primary care information needs: the difficulty of interpreting user queries, incomplete coverage of primary care concepts, intervocabulary mapping difficulties, and inconsistencies within the Metathesaurus. Experiments in automated mapping of free-text user input into a controlled vocabulary, such as MeSH, demonstrate that lexical-based approaches are useful but not perfect. Thus, lexical-based methods employing the UMLS may be helpful in capture and interpretation of user queries from primary care.

Although studies suggest that individual and combined UMLS source vocabularies incompletely cover the entire primary care domain, other evidence suggests that the UMLS is a valuable resource in representing information needs. Recent work demonstrates that progress in vocabulary mapping and answer resource identification is possible but difficult and that exploration of new methods should continue. The NLM has provided an online bibliography of health informatics research projects utilizing the UMLS.

Other Internet-based Decision Support Models in Primary Care

The emergence of the Internet and the Web has facilitated development of new models capable of providing varying levels of decision support in primary care. In 1995, Detmer and Shortliffe proposed a model of clinical query management that supports integration of biomedical information resources through a Web-based interface. The model architecture contains a Web browser, a Web server, a common-gateway-interface mediator, a representation of medi-
ical concepts (UMLS), and information resources accessible over the Internet. User queries submitted via a Web form are processed by the CGI mediator in six stages: syntactic processing, semantic analysis, selection of information resources, translation to queries, process management, and display management. They developed an application named WebMedline, which retrieves MEDLINE citations and integrates them with critical reviews published in the ACP Journal Club. More recently, Detmer and Shortliffe developed a system called MedWeaver, which integrates diagnostic decision support from DXplain, literature searching from WebMedline, and retrieval of Web sites from CliniWeb. MedWeaver employs an interface manager, query formulator, and retrieval manager to abstract the user from having to deal with separate interfaces for each resource. In a “typical” MedWeaver session, clinical findings entered by a user are processed by DXplain (through MedWeaver) to produce a differential diagnosis. From the ranked list of returned diagnoses, MedWeaver provides links for each diagnosis that show disease profile information, explain why the diagnosis appears on the list, perform literature searches, and list clinically relevant Internet sites. By providing links to individual resources, user needs are anticipated and satisfied through an “implicit model of clinician’s information needs” embedded in the interface manager.

In 1997, the Stanford Health Information Network for Education (SHINE) model outlined support for clinical decision making by unifying core health care resources in an intuitive interface over the Internet. A stated objective was to “integrate and deliver high-quality medical knowledge and the expertise of academic professionals to community-based primary care physicians.” The model also outlined methods for remote clinical teleconsultation and for awarding continuing-medical-education credit to users of the system. SHINE employs a client-server architecture and is accessible via a Web interface. Although implementation-level details were not published, the authors reported a “working prototype” that integrates knowledge from a biomedical textbook, bibliographic citation database, decision support system, practice guidelines, and primary care teaching modules.

Unlike the relatively broad-based general information needs addressable by systems like MedWeaver and SHINE, other groups are using the Internet to meet a narrower set of information needs. For example, researchers at the Mayo Clinic have used the Web to study patients with rare diseases. This approach effectively ties together otherwise isolated cases with the experts involved in highly specific areas of care. Another Web-based system, called the Physicians Research Network, provides efficient clinical trial protocol distribution and eligibility inquiries.

In summary, these and other models have begun to explore use of the Internet as a mechanism to bridge the gap between care providers in need of information and distributed answer resources. These efforts are relevant to primary care because of the broad level of service they provide. No conclusions can be drawn about these approaches, because they have not been formally tested. However, one observation seems appropriate. Although these models may differ in terms of the resources they include or who the target users might be, they share the need for a common, organizing framework that represents the information needs they attempt to address. This concern was emphasized by the developers of MedWeaver, who suggest that “a preferred approach would be to develop an external, shareable model of information needs . . .”

Discussion

This review focuses on studies of the information needs of primary care (and other) clinicians, available resources, previous approaches, and barriers to use of information resources. The authors assert that modern academic health care centers may be able to satisfy many information needs in primary care by providing Internet-mediated access to their electronic and human information resources, and we propose a model for doing so. Providing Internet-based decision support in primary care will involve experimentation in at least three key areas. First, it is necessary to develop a set of core question templates representing the types of unmet information needs that occur in primary care. At present, no readily useful and comprehensive classification of information needs exists for this purpose. Although the taxonomy should ideally be derived empirically, existing classifications might serve as a good starting point for further refinement. The eventual goal is to match information needs from the taxonomy with answer resources defined by the UMLS ISM or equivalent resource.

A second area of investigation involves assembling and managing a set of high-quality information resources to address needs specified in the taxonomy. Availability of information resources—health science libraries, bibliographic and clinical computer software applications, Web-based information, telemedicine systems, and human consultants—will vary across institutions.
Finally, the third area of investigation involves strategies for dynamically and judiciously linking questions with answer resources. For efficiency, linkages should be established in a quasi-automated manner, with minimal human intervention required. Whether done by computer or by humans, resource selection will become more difficult as the number of available resources multiplies. Unfortunately, it is not clear how existing methods of automated resource selection will scale over time.

Each advance in technology produces new ways of sharing and using information but at the same time burdens end users with the task of staying current and knowing which resources are best suited for a given need. Equally challenging is the task of evaluating existing information resources in light of newer resources, which may be more useful but also more costly. Indeed, management of these information resources in digital form has been heralded as one of the “grand challenges” in health care informatics.368

State-of-the-art technology is not ideal for addressing all information needs in primary care at this time, but the pieces of an eventual solution are coming together as continual advances in technology provide fertile ground for development of more sophisticated information systems. Triage to human resources (librarians, case managers, clinicians who screen and forward e-mail questions to subspecialists) of information requests that do not seem to map well to electronic resources may provide adequate backstopping capabilities until technology advances. The authors believe that, in the next decade, academic health care centers that leverage their resources to provide valuable information services among regional networks in primary care will probably gain a competitive advantage in the marketplace.

The authors thank the reviewers for their insightful comments and Nunzia B. Giuse for her thoughtful suggestions.

References

25. Woolf SH, Benson DA. The medical information needs of internists and pediatricians at an academic medical center.


121. Helliwell PS, Wright V. Referrals to rheumatology: a better way to reduce rates? BMJ. 1993;307:1467–70.


180. Corn M, Johnson FE. Connecting the health sciences community to the Internet: the NLM/NSF grant program. Bull
67. Jadad AR, Gagliardi A. Rating health information on the Internet: navigating to knowledge or to Babel? JAMA.


