Abstract
Objective: To identify some of the challenges that medical residents face in addressing their information needs in an inpatient setting, by examining how voice capture in natural language of clinical questions fits into workflow, and by characterizing the focus, format, and semantic content and complexity of their questions.

Design: Internal medicine residents captured information needs on a digital recorder while on a hospital inpatient service and then participated in semi-structured interviews.

Measurements: Interviews were analyzed to identify emergent themes. Recorded questions were analyzed for focus (diagnosis, treatment, or epidemiology) and format, either foreground (specific knowledge relating to an individual patient) or background (general knowledge about a condition). Semantic concepts and types were identified using MetaMap (UMLS - Unified Medical Language System) and manually.

Results: Voice recording of questions appeared to unmask residents' latent information needs. Although residents were able to record questions during workflow, there was a delay from the time questions materialized to when they were recorded. Question focus was distributed among diagnosis (32%), treatment (40%), and epidemiology (28%), and the majority of questions were background (69%). Questions were semantically complex; foreground and background questions averaged 12.6 (SD 6.0) and 9.1 (SD 6.0) UMLS concepts, respectively. MetaMap failed to recognize concepts when residents used acronyms or abbreviations or omitted key terms.

Conclusions: We found that it is feasible for residents to capture their clinical questions in natural language during workflow and that recording questions may prompt awareness of previously unrecognized information needs. However, the semantic complexity of typical questions and mapping failures due to residents’ use of acronyms and abbreviations present challenges to machine-based extraction of semantic content.

Introduction
The explosive growth of health information has made it difficult for physicians to identify and locate clinically relevant information, required for providing state-of-the-art care, during workflow. One approach to delivering information to busy physicians at the point of care is to develop an automated system where clinicians pose questions in natural spoken language via voice capture during workflow and receive answers generated through automated searching, enabling immediate use in diagnosis, management, and treatment. The current study seeks to explore two issues that will inform the development of such a system. Specifically, we wish to understand the potential barriers and solutions for capture of information needs at the point of care as well as characterize the types and semantic content of questions being asked.

Background
Information Needs of Clinicians
Integral to the success of evidence-based practice is the timely acquisition of useful and relevant information during workflow. In an era of unparalleled explicit knowledge about the causes and effective treatments of disease, and outstanding electronic resources to access such knowledge, physicians continue to have difficulty in answering clinical questions that arise when caring for patients, largely because of lack of time and insufficient training in the use of electronic databases. Interns and residents are particularly vulnerable to gaps in knowledge, given their training status, especially in the hospital setting “after hours” when consultations with colleagues are less readily available. Many residents’ clinical questions remain unanswered even though “looking things up” clearly influences their clinical decision-making.
Delivering Information to the Point of Care

Given the current time demands on typical residents and the rapidly expanding medical knowledge base, a long-term goal might include development of systems of information retrieval and delivery that originate and terminate at the point of care. An early approach was to make a computer workstation available to the staff. As the technology advanced, an “evidence-cart,” with access to information resources such as MEDLINE, followed clinical rounds to the bedside. The cart has since been miniaturized in the form of a hand-held device using wireless technology to provide access to residents to search online databases.

Enabling residents to search at the point of care does not circumvent two well-known obstacles to finding information on a timely basis. Given that an effective search often requires 10–30 minutes to complete, the time available to perform a search seems to be a rate-limiting step. As long as residents must interrupt workflow to find information, there will likely be continued under-use of available resources that could provide answers to information needs. Second, physicians have difficulty in achieving their information seeking goals and many lack the skills of effective and efficient information retrieval.

One solution to providing information to busy residents at the point of care is to create an automated system which obviates the need for residents to perform their own searches. Our Context-Initiated Question Response (CIQR) project seeks to create a question-information retrieval system composed of three steps: residents ask questions in natural language which are communicated to a central server; a search strategy is generated based on the semantic content and intent of the questions; and relevant documents containing an “answer” are returned to the residents in a timely manner.

Capturing Questions in an Inpatient Setting

The purpose of the current study was to explore several issues related to the first step of this proposed automated information retrieval system: capture of clinical questions in natural language at the point of care. Would residents be able to communicate questions in natural language via voice capture while in an inpatient setting? Inasmuch as there are no prior studies of voice capture of clinical questions, to our knowledge, we sought to determine if this form of communication was feasible in a busy clinical setting.

Characterization of Information Needs

A second goal of our study was to characterize the types of questions asked by the residents. Building a successful question-information retrieval system requires identifying the types of questions asked to direct an automated search. Prior studies suggest that the focus and format of questions may be useful in identifying target resources of a search.

The focus usually falls into one of four major categories: diagnosis, treatment, management, and epidemiology (which includes prevalence and incidence, etiology, causation or association, risk factors, disease agents, genetics, and course or prognosis). Haynes and co-workers developed “filters” that would enable human searchers to explore a reduced search space for answers to questions with a particular focus, the “clinical queries” option. A second characteristic is the format (background or foreground).

Background questions seek general knowledge and are often posed by those, such as residents in training, unfamiliar with an area and are usually answered in textbook level resources. Foreground questions seek specific knowledge relating to the circumstances of an individual patient and are usually more complex and more likely to be answered using MEDLINE or PubMed.

Semantic Content

A final goal of our study was to characterize the semantic content of the residents’ questions. An automated search strategy triggered by a question is based on an understanding of the meaning and purpose of the question. An accepted strategy is to use the Unified Medical Language System (UMLS) concepts present in the question to reflect semantic content. One method designed to match the results of a search with the original question is to use conceptual graph matching, based on semantic types and their relationships.

Our current study explores issues that will inform the development of an automatic question-information retrieval system: the feasibility of voice capture of clinical questions during workflow, and the types, semantic content, and complexity of those questions being posed.

Methods

Selection of Participants

This study, approved by the Institutional Review Board (IRB), was conducted at Columbia University Medical Center and the New York Presbyterian Hospital. Approximately 150 residents in internal medicine were invited to participate in this study through e-mail (with the permission of the Program Director). Six of them, who were rotating on the inpatient services, responded and agreed to participate. At the time of their participation, the residents were rotating either on general internal medicine, cardiology, or night float. Informed consent was obtained from each resident, each of whom was compensated 25 U.S. dollars for participating.

Collection of Clinical Questions

Residents were provided with hand-held digital recorders (Olympus DS-40), which yielded audio quality superior to cell phones, and instructed on their use. Residents were asked to carry the recorder for periods ranging from several days to several weeks. They were given no instructions as to the number of questions to record other than to record as many questions as they desired and as close to the time that the question materialized as possible. They did not receive reminders to ask questions. They were told to contact us after they had recorded 10–20 questions so we could exchange the recorders, enabling them to continue to log questions while we analyzed the previously collected batch of questions.

Interviews of Residents

After returning the recording device residents participated in semi-structured interviews that addressed a range of issues related to the experience and challenges of making recordings in these settings. Interviews lasted approximately one hour and were conducted in the departmental offices by one or two of the authors. Interviews were semi-structured and followed a defined set of questions.
Classical Content and Complexity
The authors identified UMLS concepts and terms from the questions using three computer applications: MetaMap (MMTX), the OVID MEDLINE indexing engine, and ADAM. After parsing questions with these engines, one of the analyses the authors was interested in the total number of illness-related concepts mentioned in a single question and added the number of times any one of the following illness-related semantic types were mentioned: Congenital Abnor-

Answers
As an additional incentive for the residents to participate, we searched and returned answers to the residents after they had completed their interviews (but not while they were still recording questions). A full description of the answering process, which was not a focus of this study, will be provided in a subsequent report.

Statistics
Quantitative results were compared using Student’s t-test.

Results
Voice Capture in Natural Language in an Inpatient Setting

1. Information Needs
Residents stated that having the recorder seemed to prompt them to focus on their information needs resulting in their asking more questions. One resident said: “What’s exciting about the device is that it makes me think . . . (we’re constantly thinking about questions) it really makes me acutely aware of all the things I don’t know . . . it makes me feel a lot more academic because I feel like I’m constantly thinking about the process of learning . . .”. Several residents noted that had answers been provided within a day or two they might have recorded even more questions than they did. Without answers being returned, however, they stated that there was diminished motivation to record questions. We asked residents what would have happened to their questions if they had not been able to record them. Some said that they would have written down the question and, time permitting, sought an answer later in the day or at home. Others said that given the demands on their time there would have been no reason to write down the question because answers to the question would not have been pursued. All residents reported that they recorded questions only after having gone through some prerecording information seeking or analysis (filtering). If the particular question focused on drug dosing or side effects, for example, residents chose to consult the hand-held Epocrates before recording the question. Residents also consulted the small book Pocket Medicine to see if they could find an answer before deciding whether to record a clinical question. Residents chose not to record questions requiring an urgent answer, knowing that answers were not forthcoming, and instead answered those questions immediately using an available resource.

2. Integration into Workflow
Residents found it difficult to ask their questions immediately after the question materialized. They were reluctant to speak out loud (record) in front of colleagues.
during rounds or at the nurses’ station, and had to find suitable locations to record their questions. There was thus a delay from the time a question arose to when it was recorded. At one end of the spectrum, the delay was only a few minutes. On these occasions residents found ways to interrupt workflow and locate a private place, perhaps a hallway, to ask their questions. At the other end of the spectrum, residents were unable to record questions until later in the day when they were charting their notes and preparing to leave. Some residents found opportunities to record a question en route from one location in the hospital to another.

There were also issues relating to the actual mechanics of storing and operating the recorder. Despite its relatively small size, not much larger than a typical cell phone, there were challenges in finding a convenient location to store and have ready access to the recorder. Several residents found that the addition of “yet another electronic device” (in addition to their beeper, cell phone, PDA) was actually a non-trivial matter and required adjustment during the day with the potential for interfering with workflow. Of the six residents, one was ultimately unable to use the digital recorder and thus did not record any questions.

Question Format and Focus

Sixty-five recordings were captured and subsequently transcribed using a combination of speech recognition software and human effort. Over two-thirds of the recordings were background questions composed of fewer sentences and words than foreground questions (Table 1). The former averaged 1.24 (SD 0.57) sentences and 19.0 (SD 15.6) words per question while the latter averaged 1.90 (SD 0.64) sentences and 37.2 (SD 18.0) words (p < 0.001) per question. Seventy-five percent of foreground questions consisted of more than one sentence compared to 18% of background. In a typical foreground question the resident recorded a sentence or two providing context and patient’s specific circumstances followed by a specific question.

The focus of the questions was distributed among diagnosis, treatment and epidemiology (which includes prevalence and incidence, etiology, causation or association, risk factors, disease agents, genetics, and course or prognosis). In 18 of the recordings, the focus involved a pharmaceutical agent. Residents sometimes focused on more than one issue in a single recording (15 out of 65 recordings). For example, in the question “how frequently does sarcoidosis affect the liver, how does it present, what are the therapeutic options, and how is it diagnosed?” there are four points of focus. There were thus more total points of focus than recordings (80 and 65, respectively) (Table 1).

### Table 1 - General Characteristics of Recorded Questions

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Breakdown (n, % Total)</th>
</tr>
</thead>
</table>
| Format (n = 65 recordings) | foreground (n = 20; 31%)  
|                      | background (n = 45; 69%) |
| Focus (n = 80 points of focus) | diagnosis (n = 25; 32%)  
|                      | treatment (n = 32; 40%)  
|                      | epidemiology (n = 23; 28%) |

Semantic Content

1. **Unified Medical Language System Semantic Concepts and Types**

There were 589 concepts identified in the 65 recordings, of which 360 concepts were unique. The average number of total concepts (including repeat terms) per recording was 9.06 (SD 5.98) and of unique terms was 8.12 (SD 4.57); thus, 0.94 (SD 1.76) terms were repeats. The number of concepts per question varied considerably, from as few as 3 to as many as 31. Consider the following question with 14 identifiable UMLS terms (each UMLS term is in courier typeface and underlined for clarity): “We have an HIV AIDS patient with progressive dementia as well as JC virus positive CSF. What is the definitive way to diagnose PML in AIDS patients and what type of treatment modalities are available besides HAART?”

The concepts identified in the recording were represented by 50 semantic types (Table 2, upper panel) the most frequent of which were disease or syndrome and functional concept followed by qualitative concept. In over half of the recordings there was mention of two or more illness-related semantic types and some questions contained as many as four (see question above). Most concepts of the functional concept type represented elements relating to diagnosis, treatment, disease or evidence. The category was quite broad, however, and included terms as dissimilar as pacemaker setting or CSF pressure.

2. **Semantic Complexity of Different Types of Questions**

There was no significant difference in the number of concepts or terms among different categories of question focus (diagnosis, treatment, and epidemiology). There were, however, significantly more total and unique semantic concepts in foreground questions than in background questions (Table 2, lower panel). There were also more repeat terms in foreground questions than background. Foreground questions had, on average, one additional illness-related (see Methods) or findings semantic type than background questions. Terms that provide context and nuance to the patient’s medical condition, such as those types represented in Functional Concept, Qualitative Concept, or Patient or Disabled Group, accounted for an additional two terms in foreground question (compared to background).

A majority of the recordings (75%) contained one or more of the UMLS semantic relationships that establish codependency. For example, one resident asked: “We have a patient with decompensated Hep C cirrhosis as well as severe hepatic encephalopathy that’s been refractory to lactulose. What is the utility and are there any studies that support the addition of Refiximim for hepatic encephalopathy?” This question contains several diseases that are related by (associated with), (co-occurs with), and (occurs in), and two drugs that are being compared (compared to).

Identifying Intended Semantic Content

Of the 589 concepts identified in the questions, 66 had to be inferred during manual review. Reasons why terms were not recognized by the MetaMap or the Ovid MEDLINE engines, summarized in Table 3, include use of an incorrect term or the vernacular, failure to use a preferred term, or
omitting words which, had they been included, would have resulted in successful mapping. The most challenging inferences occurred when residents implied the presence of a condition by reporting the patient’s findings or symptoms, rather than stating explicitly the condition’s name.

Abbreviations also posed a challenge to extraction of semantic content. The 65 recordings included 37 abbreviations, 10 of which neither MetaMap nor the Ovid MEDLINE indexing engine was able to recognize. Of those abbreviations recognized, often more than one concept or term was identified, only one of which was the intended term. Multiple terms were usually of different semantic types and representing differing semantic content. Acronyms based on abbreviations, such as “mersa” (when intending to communicate “MERSA”) were particularly challenging. Although Google successfully decoded “mersa”, neither MetaMap nor the MeSH indexing engine was able to. A more challenging example is HAART which, when uttered as “härt”, could refer to heart, HAART, or even hart. HAART, when entered as such, is correctly identified by MetaMap.

Discussion

Satisfying the information needs of trainees remains a considerable challenge, despite the ready availability of electronic resources. The authors approach to closing this information gap is to develop an automated question-information retrieval system that begins with residents posing questions in natural language at the point of care in an inpatient setting. The rationale of using voice capture of questions, rather than written communication, was to minimize the effort to transmit questions by eliminating the need to type, thus reducing the time it took to communicate the question; words are spoken at a significantly higher rate than either typing or texting. Inasmuch as this is the first study of voice capture in natural language of clinical questions (to the best of our knowledge), there were several issues that needed to be explored: how does voice capture fit into workflow and to what extent does it influence recognition and communication of information needs? What are the focus, format, and semantic complexity of their questions and would these characterizations differ significantly from questions captured by other means (surveys, self-reporting, interviews, or observation)?

Voice Capture during Workflow

Our interviews suggest that participation on the project and recording questions seemed to make residents aware of previously unrecognized information needs. This finding is similar to that of Ebell and White,43 who reported that physicians were more aware of information needs when prompted through interview than through self-reporting. Residents also reported that having the recording device enabled them to record questions that might otherwise have been forgotten. Green and colleagues7 found that residents reported forgetting to look things up as one explanation of their failure to satisfy their information needs. If a communication artifact (here, a recorder) in fact unmasks latent information needs and logs them so they are not forgotten, then such a system

Table 2 ▪ Number of Semantic Types and Concepts in Residents’ Questions

Most Common UMLS Semantic Types (n, % Total)

<table>
<thead>
<tr>
<th>Disease or Syndrome (n = 90; 15%)</th>
<th>Finding (N = 16; 3%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional concept (n = 90; 15%)</td>
<td>Health Care activity (n = 15; 3%)</td>
</tr>
<tr>
<td>Qualitative concept (n = 59; 10%)</td>
<td>Idea or concept (n = 13; 2%)</td>
</tr>
<tr>
<td>Patient or Disabled Group (n = 46; 8%)</td>
<td>Intellectual product (n = 10; 2%)</td>
</tr>
<tr>
<td>Quantitative concept (n = 35; 6%)</td>
<td>Laboratory procedure (n = 17; 2%)</td>
</tr>
<tr>
<td>Therapeutic or Preventive procedure (n = 35; 6%)</td>
<td>Pathological function (n = 9; 2%)</td>
</tr>
<tr>
<td>Pharmacological substance (n = 27; 5%)</td>
<td>Diagnostic procedure (n = 8; 1%)</td>
</tr>
<tr>
<td>Temporal concept (n = 26; 4%)</td>
<td>Antibiotic (n = 8; 1%)</td>
</tr>
</tbody>
</table>

UMLS Concepts per Question

<table>
<thead>
<tr>
<th>Concepts</th>
<th>All Questions (n = 65)</th>
<th>Foreground (n = 20)</th>
<th>Background (n = 45)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>9.06 (SD 5.98)</td>
<td>12.85 (SD 5.99)</td>
<td>7.38 (SD 5.21)</td>
</tr>
<tr>
<td>Unique</td>
<td>8.12 (SD 4.57)</td>
<td>11.20 (SD 4.75)</td>
<td>6.76 (SD 2.97)</td>
</tr>
<tr>
<td>Repeat (total-unique)</td>
<td>0.94 (SD 1.76)</td>
<td>1.65 (SD 1.50)</td>
<td>0.62 (SD 1.63)</td>
</tr>
</tbody>
</table>

Upper panel: UMLS semantic types identified in recordings. Total number of concepts = 549; Types appearing fewer than eight times overall are not listed. Lower panel: Number of semantic concepts in foreground and background questions. Values represent the average number of concepts (SD) per question. Mean values of foreground and background questions are significantly different (p < 0.01).

Table 3 ▪ Reasons for Failure of Machine-based Extraction of Semantic Terms

<table>
<thead>
<tr>
<th>Reason for Failure</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrect term used</td>
<td>“Basilar Angiomatosis” rather bacillary angiomatosis</td>
</tr>
<tr>
<td>Use of vernacular</td>
<td>“bugs” instead of infectious agent</td>
</tr>
<tr>
<td>Preferred term not used</td>
<td>“give” a drug rather than prescribe</td>
</tr>
<tr>
<td>Omitted words</td>
<td>“following levels” rather than following (drug) levels</td>
</tr>
<tr>
<td>Implied a condition</td>
<td>A patient with a “creatinine of 1.6” rather than kidney disease “low ejection fraction” rather than congestive heart failure</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>“Hep C” unrecognized (Disease or Syndrome)</td>
</tr>
<tr>
<td>Acronyms</td>
<td>“HAART” could be interpreted as “heart” “MERSA” unrecognized</td>
</tr>
</tbody>
</table>

Omitted words “following levels” rather than prescribing (drug) levels.
has the potential to enhance the practice of evidence-based medicine.

Residents were not able, however, to log their questions exactly when the question arose, requiring postponement of recording from as brief as only a few minutes to as long as much later in the day. The residents reported that they delayed recording to avoid speaking out loud in front of colleagues (including supervisors) or patients. Residents thus either left the nurses’ station area to go to a private location, which was not possible during rounds, or waited until later in the day to record questions. This observation is of concern because the presumed benefit of voice capture, namely instantaneous capture of critical information needs, would be diminished by the delay. Although use of more discrete voice technologies, such as Vocera, might minimize residents’ self-consciousness, silent means of communication, such as keyboarding on hand-held devices, might be preferred. This technology would no doubt introduce its own set of challenges such as the longer time it takes to type versus speak. Future studies comparing alternative means of communicating information to a server should resolve this issue.

Characterization of Residents’ Questions

The proportion of questions, posed by our residents, with a particular focus (diagnosis, treatment, or epidemiology) was similar to those reported previously by investigators using other methods of question collection (interviews, direct observation, self-reporting, and surveys, reviewed by Davies) and in outpatient settings. Neither setting nor method of question capture seemed to influence the general breakdown of question categories.

Regarding background or foreground format of questions, it was assumed that trainees would request more background information than foreground. While our observation that most of our residents’ questions were background is consistent with this prediction, it contrasts with previous observations. Green and colleagues and Cheng observed in their studies of trainees that most questions were foreground.

One possible explanation for this discrepancy is the method of question collection. Green’s study was conducted by interviewing residents after having seen a patient in an outpatient setting, which would likely have focused the residents’ question on an issue involving a particular patient (foreground question). Cheng’s study, which used surveys of residents in an inpatient setting, may also have prompted residents to think about specific cases resulting in foreground questions. These results echo the concern of Ebell and White, that the method by which questions are collected might influence the properties of the questions asked.

There are other explanations for the predominance of background questions, however. First, questions that required an immediate answer, perhaps more likely to be foreground, were not recorded. Second, residents were aware that answers to recorded questions would not be available for several days. This, too, may have focused them more on obtaining background information which would be informative but not essential to the immediate care of their patients. Last, the apparent digital recorder’s reminding residents to ask questions might have made them conscious of broader information needs.

Regardless of the explanation for a preponderance of background questions, our results suggest that, for our users, automated search engines might best be first directed to the so-called “summary level” resources (textbooks) which are likely to contain ready answers to background questions rather than the “studies level” resources (MEDLINE or PubMed). It remains to be seen if this observation and recommendation apply to residents at other institutions or on other services.

Identification of Semantic Content

As reported in the Results, over 10% of the UMLS concepts could not be identified by MetaMap and had to be inferred during manual extraction. The use of abbreviations, acronyms, incomplete descriptions, or the failure to articulate explicitly the presence of a condition or illness resulted in mapping gaps that would completely disable an automated search engine. A key question, which we are unable to answer from the results of this study, is whether the overall frequency and types of mapping gaps were the result of voice capture in natural language. Given that prior studies using nonvoice means of question capture did not undertake a similar semantic analysis to ours, we cannot compare results. The only certain conclusion we can draw is that homonym confusion (HAART, heart, hart) is a direct result of voice capture. However, it could be that other examples of mapping gaps are the result of voice capture. Were residents inclined to leave out key terms that they would otherwise have included in a written communication? When “thinking aloud” using the recorder, were residents more likely to describe findings and symptoms (imply the presence of a disease) rather describe the actual illness or use abbreviations?

Regardless of whether voice capture increases the frequency of mapping gaps, compared to written communication, a strategy to reduce gaps would benefit machine-based semantic extraction from questions collected by any means of capture. Residents could be trained to ask questions in a manner that reduces semantic ambiguity and facilitates easy identification of the intended semantic content. If clinicians can be taught to ask “well-built questions” (in the PICO format, see Methods section) to improve searching, perhaps they can also be taught to ask “well-formulated questions.” Clarity would be improved if they used terms from controlled terminologies (UMLS) rather than the vernacular, used as few abbreviations as necessary and as many terms from a medical dictionary as possible, avoided acronyms altogether, avoided implying concepts by articulating key words, identified the disease or condition, when possible, rather than describing a constellation of symptoms or findings, and, above all, asked one question at a time. Having to compose well-formulated questions, however, might dissuade participants from asking questions as well as increase the cognitive load.

Whether or not training residents to ask well-formulated questions is feasible and productive needs to be addressed in future studies.

Semantic Complexity

Our studies demonstrate that foreground questions are semantically complex: they contain nearly ten concepts per question, often make reference to two or more diseases, conditions, or drugs, and contain one or more of the semantic relationships...
that establish codependency. We cannot comment on whether capturing questions in natural spoken language influenced complexity because prior studies using different methods of collection (survey, self-reporting, interview, or direct observation), some of which explored an aspect of “complexity”\textsuperscript{16,19,45,51}, did not undertake a similar semantic analysis.

Semantic complexity of foreground questions might have resulted from the filtering that took place before recording questions. Residents might have found answers to less complex foreground questions with Epocrates\textsuperscript{39} or Pocket Medicine.\textsuperscript{40} Prescreening may have eliminated the simple questions such as “I have a patient with infection x. How long should I treat them with drug Y?” Or, the complexity of the foreground questions could reflect the inpatient setting on an internal medicine service. Patients in this setting are often elderly or chronically ill with numerous comorbid conditions all of which need to be considered in diagnosis and management decisions.

The observation that a considerable proportion of the questions are extremely complex may provide a more realistic view of information needs, one to be considered when designing algorithms to search resources and retrieve information to provide answers. In prior work designed to develop such search engines, investigators used questions to evaluate algorithms that were appreciably less complex than those recorded by our residents.\textsuperscript{29,52,53}

**Limitations of the Study**

Throughout the study residents were aware that answers to their questions would not be provided sufficiently quickly to influence management. Had answers been returned in a timely manner, as proposed in the CIQR project, perhaps the proportions of questions with a particular focus or format would differ from that observed in this study. A similar analysis on a second collection of questions obtained when the CIQR system is in place may resolve this issue.

The authors can not make any generalization about the complexity of foreground questions or the frequency and types of mapping challenges (abbreviations, incorrect terms, and implied terms), which ranged considerably, because questions were collected from a small group of internal medicine residents. Had questions been collected from residents on a pediatric service, whose patients are young and unlikely to have as many comorbid conditions, the questions would likely have been simpler with fewer semantic terms.\textsuperscript{54} One expects also that there is variation among residents as to the formality of the language chosen to communicate information needs. These issues, too, can only be resolved by studying additional residents in different settings and services.

**Future Work**

The potential benefits of voice capture of clinical questions, ease of use, and unmasking of information needs, will not be fully realized if the method increases the frequency of mapping gaps. Future work, comparing semantic ambiguity in questions obtained by a different modality, such as keyboarding, should resolve this issue. Also, given the semantic complexity of foreground questions, strategies are needed to reduce complexity such as ignoring repeat and context terms not related to the patients’ condition, and applying a combination of preference rules,\textsuperscript{55} semantic-type filters,\textsuperscript{32} and natural language processing.\textsuperscript{56}

**Conclusions**

Our results demonstrate that capturing questions through voice capture in natural language at the point of care is feasible and may alert residents to the full measure of their information needs, which are largely requests for background information. The potential advantages of voice capture, the ease of communication and prompting awareness of gaps in knowledge, may be offset by the disadvantage of having to speak out loud in front of colleagues, nurses, and patients. Questions tended to be semantically complex and contained content that was sometimes not expressed in a manner that allowed for successful machine-based extraction of key semantic concepts.

**References**


46. Haynes RB. Of studies, syntheses, synopses, summaries, and systems: The “5S” evolution of information services for evidence-based health care decisions. ACP J Club 2006;145(3).


