Assessing Data Relevance For Automated Generation Of A Clinical Summary

Tielman T. Van Vleck, MA¹, Daniel M. Stein, MD¹, Peter D. Stetson, MD MA¹², Stephen B. Johnson, PhD¹
¹Department of Biomedical Informatics, Columbia University, New York, NY
²Department of Medicine, Columbia University, New York, NY

ABSTRACT
Clinicians perform many tasks in their daily work requiring summarization of clinical data. However, as technology makes more data available, the challenges of data overload become ever more significant. As interoperable data exchange between hospitals becomes more common, there is an increased need for tools to summarize information. Our goal is to develop automated tools to aid clinical data summarization. Structured interviews were conducted on physicians to identify information from an electronic health record they considered relevant to explaining the patient’s medical history. Desirable data types were systematically evaluated using qualitative and quantitative analysis to assess data categories and patterns of data use. We report here on the implications of these results for the design of automated tools for summarization of patient history.

INTRODUCTION
A fundamental task in the practice of medicine is to sort through large amounts of data to find pertinent clinical information. This challenge becomes increasingly significant as advances in diagnostic technologies and the biosciences augment the traditional set of data available to clinicians. Much attention has been given to the notion of evidence-based medicine, and how to address the information needs of clinicians to answer clinical questions and support decision-making.¹² Less attention, however, has been focused on how to assist physicians in navigating the ever-growing amount of clinical data that is accumulating for each individual patient.

Modern electronic health records (EHR’s) provide a portal into a vast information space for each patient and allow data from a variety of sources to be aggregated into one place. EHR’s amass information from local data silos including laboratory data, imaging studies and clinical notes. Furthermore, recent efforts to connect institutions via regional health information organizations (RHIO’s) have the potential to expand available resources to include data from outside healthcare providers and facilities.³ This rich information source has the potential to enhance clinicians’ ability to efficiently gather patient data.⁴ However, there is also rising potential for information overload – that is, clinicians may have access to so much information that it becomes difficult to differentiate that which is pertinent from that which is noise.

At each patient encounter, a physician must review and process previously documented patient history in an ever-shortening period of time. Any system that could increase efficiency of this task would not only improve workflow efficiency but also has the potential to improve outcomes for patient care. Previous work in this area has suggested feasibility and benefits of an automated patient summary sheet for outpatient visits; patients for whom such sheets were used demonstrated improved health outcomes.⁵ These early studies demonstrate the potential utility of, and need for, automated generation of an electronic clinical document that could efficiently summarize key pieces of information. We believe that in order to design such a system and ensure its viability and maximal utility, it is essential to have a deeper understanding of what information is of most importance to physicians when reviewing a patient’s medical record. The goal of this study was to capture and characterize the pieces of information physicians consider relevant when developing a comprehensive cognitive representation of a patient. We also attempted to determine where in the medical record these pieces of information are found. This insight should facilitate the creation of automated systems to provide concise and highly useful patient summaries for a wide variety of clinical uses.

METHODS
Overview: A series of structured interviews were performed on five Department of Medicine residents at New York-Presbyterian Hospital to identify phrases in the medical record that each physician perceived to be relevant for describing the patient’s history. Subjects were provided with complete, de-identified patient records for three common general medical admissions. For each case, the subject was asked to acquaint him/herself with the patient and then to underline information critical to describing the patient to a colleague.
Primary data sources for physician review of the selected cases included the complete medical record including all typical clinical data types (e.g. lab, radiology, medications, ancillary reports and provider notes).

Primary measures included the number and information classes of phrases identified across different divisions of the medical record.

Clinical Document Collection: A query was performed on the clinical data warehouse at New York-Presbyterian Hospital (NYPH) to identify three admissions of patients admitted to the hospitalist service. Inclusion criteria were an admission diagnosis of one of three common conditions: congestive heart failure, pneumonia, or any of the common sequelae of diabetes. These primary admissions were considered only if they had a preceding hospital admission and a subsequent admission and/or clinic visit. Further inclusion criteria included: a primary admission of at least four days in duration, a discharge summary for the primary admission, and at least one follow-up note written during the subsequent admission or clinic visit. Once patient admissions matching the three primary diagnoses had been identified, related patient documents were gathered from our clinical data repository.

All patient data was de-identified prior to the structured interviews with study subjects. All data collection and analysis was completed in accordance with Internal Review Board procedures and approvals.

Study Subjects: Five physicians were recruited for the study. All were second and third year residents in Columbia University Medical Center’s Department of Medicine.

Physician Identification of Relevant Historical Information: Each subject was given a complete set of de-identified notes from three sample patient admissions, excluding the discharge summary. Discharge summaries were omitted in order to encourage subjects to use primary data sources as discharge summaries tend to be an imperfect review of existing information. The order of the patients presented was varied to avoid bias as the physicians became familiar with the task. Subjects were asked to imagine types of information that they predicted to be pertinent in the future and to create a list of these categories, assigning a pen color to each one. They then reviewed each admission and considered what information in the paper chart would be relevant to them or a colleague during a subsequent hospitalization or clinic visit.

After gaining a brief understanding of the patient, subjects were asked to review the entire record, underlining relevant phrases using the colored pen associated with the appropriate category. They were reminded that this task would be similar to identifying information for inclusion in a discharge summary, except that it should not include all information but should instead be filtered to include only differentiating information that would be relevant to future evaluation.

Interviews were recorded to capture the order of tasks performed. Subjects were encouraged to follow a think-aloud protocol, explaining when they were switching from one section to another and explain any notable concepts identified in the text.

Data Analysis: Underlined text was transcribed with the assigned category, along with associated division of the record, document type, and section of the document as appropriate. Categories were grouped into logical sets for comparison across subjects. Data were then analyzed to identify the most common sources of relevant phrases.

RESULTS

During the structured interview process, physicians underlined 824 phrases that they considered relevant to explaining the patient’s history. The number of terms selected varied significantly between subjects, ranging from 60 to 327. Subjects developed a list of categories considered relevant to the process of patient familiarization and applied one to each phrase of interest. The categorization process had varied results, with the number of categories created ranging from eight to fourteen as listed in Table 1. After the interview process, these categories were manually consolidated into eight groups including: Labs and Tests, Problem & Treatment, History, Findings, Allergies, Meds, Plan and Identifying Info. While no consolidation perfectly accommodated all categories identified, this was required to facilitate summarization of the results such as in Figure 1.

![Figure 1: Phrases identified for each category group](image_url)
Phrases identified were consolidated by: EHR division, note-type or lab description, clinical note section, and category. The top five divisions, tests or notes and note sections are summarized in Table 2.

<table>
<thead>
<tr>
<th>Divisions</th>
<th>Tests/Notes</th>
<th>Note Sections</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Notes</td>
<td>Initial Visit Note</td>
<td>1. CC/HPI</td>
</tr>
<tr>
<td>2. Laboratory</td>
<td>Progress Note (Intermediate)</td>
<td>2. Data/Lab</td>
</tr>
<tr>
<td>3. Cardiology</td>
<td>Cardiac</td>
<td>3. Physical Exam</td>
</tr>
<tr>
<td>4. Radiology</td>
<td>Catheterization</td>
<td>4. Assessment</td>
</tr>
<tr>
<td>5. Diagnoses</td>
<td>Progress Note (Final)</td>
<td>5. Plan</td>
</tr>
</tbody>
</table>

Table 2: Top five citation sources in three categories

Each subject had a very different approach to the defined task. Some were careful to include only phrases critical to future analysis, while others did not filter concepts as strictly. Notes were by far the most popular source, but while some subjects drew almost entirely from notes, others relied much more heavily on laboratory data. Further analysis compared use of different sections of the EHR by disease, as presented in Figure 2.

Figure 3 examines where in the EHR output different categories of data were identified.

The path followed by each subject varied little. In every instance, the resident first located the admission note and read it thoroughly. They then skimmed through the intermediate progress notes until they got to the last progress note. All explained that they would have referenced the discharge summary if it had been available, but that the last progress note was nearly as good for providing a final opinion on the patient course.

During the interview, several residents expressed significant frustration over duplicate information copied and pasted from one record to the next. While this problem is well known and documented, the frustration expressed by the subjects provided further proof of the need for some type of solution.

After the interview, one subject voluntarily drew up a hypothetical clinical summary of the patient. This was done in a very problem-oriented format.

**DISCUSSION**

The results of this study demonstrate that through the use of a cognitive walkthrough methodology, we were able to capture concepts relevant to a physician in the context of familiarizing him or herself with a patient’s medical history. We were successful in capturing categories of clinical data that would facilitate this task. We captured 824 phrases relevant...
to summarizing the patient, as well as 62 categories developed by the five subjects for grouping these phrases. The categories were consolidated into eight logical groups that could serve as a platform for structuring an initial demonstration of a clinical summary document.

Categories created by physicians largely mimic the traditional categories of the medical record which are thoroughly taught in medical school. Despite asking subjects to avoid these traditional categorizations, all were inclined to resort to these groupings. Whether or not these are the most efficient for patient summarization, it appears clear that every physician follows this methodology and hence future work summarizing patient history should consider this structure as it is core to the clinical thought process.

Much work has been done evaluating the cognitive process of physicians during the process of patient evaluation, mapping mental models constructed during the process of diagnosis. As the goal of this research is to build a tool that will assist constructing a mental model of the patient, it is not necessary to study the cognitive process of learning critical information or mental representations of this information, but instead analyze criteria of information physicians consider worthy of inclusion. This study attempts to identify criteria for such information that can be used to programatically identify key information in narrative text that is relevant to a given patient scenario.

The greatest challenge in this study proved to be regulating the scope of the results selected by our research subjects. The number of terms selected varied by a factor of five. For the hypothetical task, it is possible, though unlikely, that under similar conditions one physician would analyze five times as many data points as another.

A major limitation of the study was the paper-based nature of the procedure. At times, subjects expressed a desire to access information not included in the paper report such as Pulmonary Function Tests, EKG’s, various consult notes not included and weights and vital signs stored in a separate system used in the hospital. Subjects also found that without the visualization tools built into the EHR at NYPH, such as those that allow one to easily visualize lab result trends, they performed the requested task differently. Two subjects remarked that they would have relied extensively on laboratory data had trend charts been available, but with simple tables of lab data they relied more on other physicians’ interpretations of the results. Future tests should probably address this disparity, and try to more closely mimic the everyday working conditions of the physician.

By not including the discharge summary in the survey, subjects were forced to do a more thorough analysis, however future studies must review discharge summaries in some capacity to account for what information was considered novel in the summary as well as what was omitted.

Implications: For physicians to process all relevant information on a patient, cognitive tools are essential for assisting in the rapid comprehension of large amounts of information. For quantitative information, this can be as simple as using a small chart to represent laboratory data. The task is more challenging for qualitative descriptions, and will require tools capable of aggregating text relevant to particular subjects.

This study is a formative evaluation of a methodology to identify a gold standard for use in programmatic generation of a clinical summary describing key information from a patient’s medical history. Subsequent research will look to identify inclusion criteria for the summary. Some criteria may be determined through simple rules, while others will require more complicated analysis.

It became clear through the interviews that simple rules based on location in the medical record may provide a useful start to identifying data to include in a summary. For example, every physician considers advance directives such as a “Do Not Resuscitate” order to be crucial to know about. Clearly, these should be included in any type of patient overview presentation.

However, most concepts in the clinical summary will have to be identified using more detailed techniques. Medical language understanding systems have been developed to analyze meaningful criteria of medical texts. The MedLEE natural language processor is a medical language understanding system developed at Columbia University by Dr. Carol Friedman and will be applied to these notes. This analysis could yield criteria on concepts within phrases of the text that might be indicative of relevance. Additionally, semantic indicators such as prefix words may prove to provide meaningful indication of a phrase’s relevance.

With these criteria available for phrases within the medical record, results of this study may serve as a useful gold standard in determining rules for identifying phrases of interest from the medical history. While it may be possible to manually curate rules from these data, a machine-learning (ML) algorithm may be ideally suited for identifying trends
from these results. For example, in interviews subjects generally found medications listed in the final progress note to be relevant as discharge medications. With NLP tags denoting all medications in the note, a ML engine is likely to recognize that final medication lists are relevant to future work, while intermediate medications not included in the final note are less important.

**Future Work:** As this research will serve as the foundation for further work developing tools for the automatic generation of a clinical summary, further analysis of these data will focus on identifying additional characteristics of the phrases subjects found relevant. The full text will be processed with MedLEE to enable analysis of semantic properties for concepts within phrases of interest. With this information we will be able to identify criteria for phrases of interest and also identify where in the overall text similar phrases were not included. This is important for analyzing the specificity of the tool, as it is critical in any summarization to cite as little information as possible to convey the intended information. NLP analysis will also provide UMLS and MeSH tagging of phrases, enabling analysis of UMLS and MeSH categories indicative of relevance.

Additional studies may be required for a more thorough evaluation asking physicians to review and summarize the entire patient history. This study intentionally kept the task simple by focusing on a single admission. This is important, as identifying information key to each admission is critical to a clinical summary. However, it would also be of great interest to ask physicians to perform a thorough analysis of the entire patient record and draw up a sample clinical summary of their own.

Several investigators have underscored the need for electronic “discharge summaries” for rapid and consistent delivery of hospitalization information to primary outpatient physicians, and potentially for facilitating the automated detection of adverse events.8-10 The procedure in this study was very similar to the task of generating a discharge summary. Results of this study could be integral to construction of tools assisting in the consolidation of information for inclusions in the discharge summary.

**CONCLUSIONS**

This study elucidated ways residents familiarize themselves with a patient using the contents of an electronic health record. The categories and patterns of data use provide a starting point for developing tools to automate the task of automated generation of a high-level, patient summary.

**ACKNOWLEDGEMENTS**

This work was supported in part by NLM training grant N01-LM07079 (TTVV, DMS), NLM grant K22-LM008805 (PDS), and RO1-LM008635 from the NLM.

**REFERENCES**


