HARVEST, a longitudinal patient record summarizer

Jamie S Hirsch[1,2], Jessica S Tanenbaum[3], Sharon Lipsky Gorman[1], Connie Liu[1], Eric Schmitz[4], Dritan Hashorva[4], Artem Ervits[4], David Vawdrey[1], Marc Sturm[4], Noémie Elhadad[1]

ABSTRACT

Objective To describe HARVEST, a novel point-of-care patient summarization and visualization tool, and to conduct a formative evaluation study to assess its effectiveness and gather feedback for iterative improvements.

Materials and methods HARVEST is a problem-based, interactive, temporal visualization of longitudinal patient records. Using scalable, distributed natural language processing and problem salience computation, the system extracts content from the patient notes and aggregates and presents information from multiple care settings. Clinical usability was assessed with physician participants using a timed, task-based chart review and questionnaire, with performance differences recorded between conditions (standard data review system and HARVEST).

Results HARVEST displays patient information longitudinally using a timeline, a problem cloud as extracted from notes, and focused access to clinical documentation. Despite lack of familiarity with HARVEST, when using a task-based evaluation, performance and time-to-task completion was maintained in patient review scenarios using HARVEST alone or the standard clinical information system at our institution. Subjects reported very high satisfaction with HARVEST and interest in using the system in their daily practice.

Discussion HARVEST is available for wide deployment at our institution. Evaluation provided informative feedback and directions for future improvements.

Conclusions HARVEST was designed to address the unmet need for clinicians at the point of care, facilitating review of essential patient information. The deployment of HARVEST in our institution allows us to study patient record summarization as an informatics intervention in a real-world setting. It also provides an opportunity to learn how clinicians use the summarizer, enabling informed interface and content iteration and optimization to improve patient care.

Key words: electronic health record; summarization; natural language processing; visualization

BACKGROUND AND SIGNIFICANCE

With increasing numbers of observations recorded about patients in their records, providers are faced with an overwhelming amount of complex, raw data points, with little time for making sense of them all.[1] This phenomenon of information overload has been observed in several care settings, from outpatient clinics to hospital admissions and the emergency department (ED).[2–4] With the advent of health information exchange, review of patient data will only become more complex and time consuming.[5,6] One of the promises of the electronic health record (EHR) is to support clinicians at the point of patient care. Clinical information systems, unfortunately, seldom provide effective cognitive support—that is, they do not present information in an optimal format to clinicians when and where it is needed.[7] The lack of usable and effective patient review capability in the EHR forces clinicians into spending precious time resources reviewing low-level, unrelated data points trying to construct a mental model of the patient that may result in delayed care and diagnostic and treatment errors.[4,7]

The need for better health information management and visualization tools has long been recognized. Powsner and Tufte,[8] in the early 1990s, proposed the graphical representation of monitoring data combined with essential information about a patient in intensive care as a way to relieve information overload. Summarization of individual longitudinal medical records, 'the act of collecting, distilling, and synthesizing patient information for the purpose of facilitating any of a wide range of clinical tasks' is an open and active field of research in informatics.[9] Important research has been undertaken on visualization of patient histories, such as the LifeLine and the KNAVE projects.[10,11]

RESEARCH AND APPLICATIONS

Summaries constructed as dashboards, collecting information from the structured part of the EHR, have also been investigated. In specific cases, studies have shown that access to summaries had a positive impact on care, ranging from disease management to new problem detection,[12–15] further motivating the need for robust summarization integrated into the EHR.

To understand how physicians make sense of a patient record, we conducted an informative study asking physicians to summarize longitudinal patient records.[16] Analysis of the physicians' summarization processes and additional interviews conducted with physicians: (i) discovered that identifying and synthesizing information is a cognitively complex and time-consuming task, even for expert users of EHRs; (ii) showed that critical summary content is conveyed in the notes, and, as such, notes are a primary source of information when creating a summary; and (iii) indicated that, while there was no established standard for organizing summary content, a problem-oriented view of the patient was a successful summary organization strategy. These findings particularly resonate with the vision of the problem-oriented record.[17–19] In this framework, these problems go beyond billing codes and existing problem lists in the EHR, and, as such, parsing of notes is a critical aspect of the summarization process. There has been much research on visualizing document content in a collection,[20,21] on which this work builds.

OBJECTIVE

We describe HARVEST, a longitudinal patient record summarization system. HARVEST is an interactive, problem-oriented temporal visualization, as extracted from a patient's record. The summarizer differs from previous work in three critical ways: (i) it extracts content from the patient notes, where key clinical information resides; (ii) it aggregates and presents information from multiple care settings, including inpatient, ambulatory, and ED encounters; and (iii) it operates on top of commercial EHR systems, and is available for all patients in our institution, not just a curated dataset or for specific patient cohorts. We followed a user-centered, iterative design, by developing an initial design, implementing and deploying it into the EHR at New York Presbyterian Hospital (NYPH), and gathering feedback in order to improve the system. We also report on a formative evaluation carried out to assess the ability of HARVEST to support physicians when reviewing patient information and to gather user feedback for future iterations.

MATERIALS AND METHODS

In this section, we describe the clinical environment, HARVEST's architecture, back-end and front-end processing, and the setup for the first of a series of evaluation studies of HARVEST.

Clinical environment and deployment sites

NYPH has six primary inpatient facilities, including two academic medical centers, and several outpatient facilities in its Ambulatory Care Network. Together they account for 127 000 hospital discharges, 1 650 000 outpatient visits and 311 000 emergency visits annually. All NYPH providers rely on iNYP (on average, 9600 unique users per month).

iNYP is a web-based results review application focusing primarily on the display of textual rather than graphical data. The core display is based on the institution's legacy WebCIS results review application.[22,23] A left-hand panel contains links to components of the chart (eg, laboratory data, clinician notes, radiology reports, and pathology reports), and detailed data are displayed in a large central panel (figure 1). A rudimentary search function exists based on document type, document author, or laboratory test type. A summary tab displays grouped laboratory data (eg, basic metabolic panel) over time in tabular format.

HARVEST was deployed into the clinical information review system at NYPH as a β release in September 2013. It is available for all patients in the institution, but clinician access has been restricted to internal medicine and emergency medicine departments thus far.

HARVEST architecture

The HARVEST architecture consists of two online processing modules: online distributed HL7 message and visit parsing for all patients as they are generated by the EHR and web-based visualization for a specific patient (figure 2). Both the back-end and front-end processing are independent of the clinical information system and rely on an HL7 feed as input and a web server as output. As such, HARVEST can function on top of any clinical system, provided that an HL7 feed is available. User authentication and patient selection is handled by the clinical information system.

Back-end scalable, distributed visit and note processing

HARVEST's back-end processing consists of the distributed parsing of visit information and indexing of clinical notes as they are generated, as well as the salience computation for problems to be visualized in the cloud.

Natural language processing (NLP) of notes

All note types in the institution, as written by all provider types (ie, physicians, students, nurses, social workers, nutritionists, etc), are processed. For each note, document structure (section boundaries and section headers, lists and paragraph boundaries, and sentence boundary detection) and mentions of problems are identified with our in-house named-entity recognizer, HealthTermFinder.[24,25] HealthTermFinder identifies concept mentions and normalizes them to semantic concepts in a given terminology. Since HARVEST is a problem-oriented visualization, we restrict the terminology to the concepts appearing in the Unified Medical Language System (UMLS) semantic group, 'Disorder',[26] and filtered to the Core Problem List Subset of Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT).[27] In this version of HARVEST, attribute identification was skipped; thus, negated and uncertain terms were still indexed.

Problem salience computation

Salience weights of problems extracted from the notes are computed dynamically to reflect both the frequency of the

Figure 1: Screenshot of iNYP, the standard clinical information review system at New York Presbyterian Hospital (NYPH) for a deidentified patient. The left frame displays the contents of the patient chart. Selecting a category opens that section in the main frame. In this example, the main frame displays the notes in chronological order with the author names, and selected note contents appear in the bottom frame.

concept in the patient notes in a given time slice of the record and the prevalence of the concept across all patients in the institution. Salience weights are based on the established term frequency inverse to document frequency (TF*IDF) framework for scoring importance of terms in a document collection.[28]

Scalable, distributed processing

To enable parsing and salience computation at scale, we created a distributed computing infrastructure using Apache Hadoop and implemented a map-reduce version of our NLP system to parse the notes from HL7 messages as they are generated, as well as to parse the visits from the visit information feed in the EHR (ie, visit dates and primary International Classification of Diseases (ICD)9 billing codes). This infrastructure enables us to accommodate the large volume of documentation generated at NYPH (650 000 notes per month). A small four-node cluster processes 20 000 notes per second compared with 500 notes per second in a non-distributed computing environment. Processed notes, extracted problems, salience weights and meta-data are stored on Apache HBase, a distributed database, such that new content is added and integrated smoothly and without downtime in the system.

Front-end visualization

The front-end processing for HARVEST is a web-based visualization. When an authenticated user invokes the summarizer for a particular patient, the database is queried for summary content and returns a JSON message containing all visit, note, and problem information. The interface visualizes the JSON messages into three panels—a timeline, a problem cloud, and

Figure 2: HARVEST system architecture. The distributed online module parses and indexes notes and visits as they are generated from the electronic health record, so that, for all patients in the institution, the HBase database stores information about visits, notes, and problems extracted from the notes. An authenticated user can invoke the online web-based visualization for a particular medical record number, as part of the clinical information review system. CUI, concept unique identifier; MRN, medical record number.
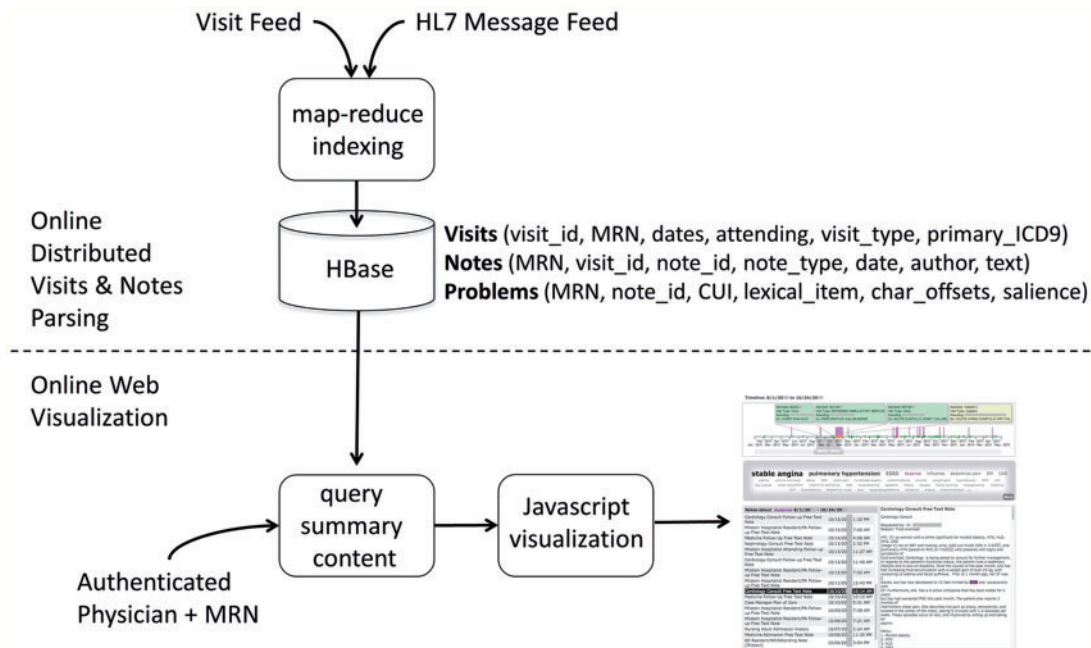
a note display panel—through HTML5 and Javascript code. The problems' salience weights determine the problems' size in the cloud. The timeline was implemented using HTML5 canvas, and the cloud and note panels were implemented with JQuery.

Formative evaluation of HARVEST functionality
The goal of this formative study was to assess the ability of HARVEST to support clinicians in reviewing patient information and to gather feedback for iterative design.

Study subjects and patient selection
This study was approved by the Columbia University Institutional Review Board. Subjects were recruited via invitation email sent to all nephrology fellows and internal medicine residents. Participation was voluntary and compensated. All subjects were experienced users of iNYP, the institution's standard data review system, but had never seen HARVEST before.

Cases for patient review were selected from the NYPH population, and selection focused on complex patients with multiple comorbidities, frequent interactions with the healthcare system, and a combination of outpatient, inpatient, and ED visits. Patients had a mean of 48.5 months of clinical follow-up at NYPH (range 9–62 months) with several medical providers,

numerous visits, and copious documentation with a sizeable number of problems and unique problems, as extracted by our NLP system (table 1). All records were checked to ensure that the physician participant had not cared for the patient at any time previously.

Study protocol
Each study subject saw four individual patient cases, two with just the use of iNYP and two with HARVEST. When in the HARVEST condition, subjects were asked to rely solely on HARVEST, but were allowed to access other iNYP sections. Subjects in groups of four were evaluated on the same four patient cases, with the order of the patients and condition alternated (figure 3). Fellows were evaluated in their own cohort of four cases and were not examined on the same cases as residents.

Each case scenario lasted 20 min, during which subjects completed a 35-item questionnaire. This time limit allowed sufficient time for questionnaire completion yet applied some element of time pressure mimicking real-world conditions (although participants were not explicitly asked to complete the questionnaires as quickly as possible).

Task items focused on date finding (eg, 'How soon after [hospital] discharge did the patient have outpatient clinical follow up?'), clinical fact or event finding (eg, 'Does this patient

Table 1: Electronic health record representation of the 12 study patients

	Mean	Median	Range
Total clinical follow-up in the hospital system (months)	48.5	57.5	9–62
In 2 years preceding study, number of:			
Healthcare visits	65.7	57.5	20–133
Provider types	12.75	11.5	4–20
Documents	266.25	226.5	75–874
Document types	62.25	56.5	12–137
Document authors	111	90.5	22–379
Problem concepts	5129.5	3877	1001–16 502
Unique problem concepts	253.75	271	94–462

Figure 3: Study protocol for groups of four subjects tested with four patient cases. This study consisted of a total of three groups of four subjects each (total 12 subjects and 12 patient cases). Light gray boxes indicate study condition with the use of HARVEST and dark gray boxes indicates use of iNYP alone. Overall, the study had 12 subjects and 12 patient cases.

have a history of rash?'), clinical comparisons (eg, 'What accounted for the change [in clinical problems over a period of time]?'), and clinical synthesis (eg, 'What were the five most prominent or important problems [over a period of time]?'). The questionnaire was identical for every patient case except for three questions, which asked specifically about a chronic medical condition for the given patient (eg, hypertension, congestive heart failure, diabetes mellitus). The final question, which was solicited once the patient record was closed, asked for a patient 'one-liner,' a summary sentence encompassing items such as age, gender, relevant or important comorbidities, and significant medical details. Overall, the task assessed ability to

extract, compare, synthesize, and recall clinical facts about a patient. At the conclusion of the study, participants were asked to complete a post-study questionnaire, with Likert-type and free text questions to ascertain perceived usability and experience with HARVEST.

Data collection

This study used Morae Recorder, V.3.3.3, data capture software, which included video, audio, and mouse and keyboard tracking. Study subjects were asked to follow a think-aloud protocol. All participants used the same computer, and had access only to iNYP/HARVEST and a Microsoft Word document

with the questionnaire. All answers were entered on the Word document, after which it was transformed into a read-only version to prevent error.

Statistical analysis

Two authors with medical knowledge independently completed an answer key for each patient case, with differences adjudicated in an open setting in order to arrive at a final consensus reference answer key. Each completed questionnaire was then compared with the reference standard for scoring. All but two questions were given a maximum of one point (and were normalized to one if there were subcomponents). The two synthesis questions were weighted more (two points each). For the sake of computing an overall score, unanswered questions were scored as incorrect. The final score for a given questionnaire was normalized to the maximum number of answerable questions for that scenario.

Sensitivity analysis was conducted by variably excluding outlier scores (defined as any value below the third quartile or above the first quartile 1.5 times the inner quartile range) and unfinished questions (due to time constraints). The one-liner question was not included as part of the overall score, but was scored separately (on a scale of one to three, with one being inadequate and three being comprehensive) and analyzed. Overall time to completion (in seconds) of each questionnaire was captured with Morae Recorder; the one-liner was not included in the timing.

Given the overall sample size and uncertain distribution, non-parametric statistical analyses were performed. Because of the repeated measurements of each subject, Wilcoxon signed-rank test was performed for comparison between two study conditions. Study participants were exposed to the iNYP and the HARVEST conditions twice, and analysis was conducted with collapse of scores (by averaging) for each study condition, allowing direct comparison between a single iNYP score and a single HARVEST score for each participant. As a loss of power occurs with collapse of features, sensitivity analysis was performed, comparing all scores from the iNYP and HARVEST conditions without any averaging. Correlations were assessed using Spearman's rank correlation, and Fisher's exact test was used for analysis of categorical outcomes.

All analyses were conducted using the R language and environment for statistical computing, V.3.1.0.

RESULTS

HARVEST summarization and visualization

HARVEST is a web-based interactive, problem-oriented, temporal visualization of a patient record. It renders patient data through a timeline, a problem cloud, and two note panels (figure 4).

The timeline expands and contracts, and a slider facilitates focus on specific time periods. Marks on the timeline indicate different types of visit—including outpatient, ED, and inpatient admissions—and information bubbles above visits provide meta-information, including visit date, attending physician, and primary billing code. The timeline enables users to assess the visit patterns of a patient (eg, a patient with frequent ED or clinic visits).

The problem cloud displays concepts extracted from the notes in the patient record for the selected time range in the timeline, advancing beyond simple billing codes or problem lists. Problems are ordered in the cloud by their salience weight for a given range. Normalization of problems goes beyond traditional word-cloud capabilities: synonyms and lexical variants of a given problem are all aggregated into each problem (see for instance, the problem 'dyspnea' and its synonym 'SOB,' or 'shortness of breath,' in figure 4). When a problem is selected, the timeline displays the different occurrences of the problem throughout the record, allowing users to assess the patterns of documentation for that particular problem.

The note panel consists of a list of notes in the patient record for the selected time range in the timeline. If a problem in the cloud is selected, the list is filtered to display only the notes that mention the selected problem. Individual notes can be viewed as well, with a selected problem highlighted throughout. This functionality enables users to access the actual context in which problems were mentioned and gather full information about the problem from existing documentation.

Formative evaluation

Twelve physicians partook in the evaluation, eight internal medicine residents (4 senior residents, 2 junior residents, and 2 interns) and four nephrology fellows (1 senior fellow and 3 junior fellows). Subjects completed four patient cases, for a total of 48 completed task questionnaires covering 12 unique patient cases across two testing conditions. Multiple exposures to the same testing condition did not improve scores, as no statistically significant intrasubject difference was seen between trial one and two for each condition (iNYP p = 0.923 and HARVEST p = 0.615) or between conditions (p = 0.850). The scores for each of the two conditions were then collapsed by averaging for each participant, yielding a total of 12 scores each for iNYP and HARVEST comparisons.

In comparison, between the two study conditions, no statistically significant difference was observed in the overall scores on the task questionnaires (p = 0.569; table 2 and figure 5). These findings did not change in various sensitivity analyses comparing all 24 scores in each condition (p = 0.546).

Additional analyses were conducted examining the effect of excluding outlier cases, incomplete questions, patient case order, and physician type (ie, fellow vs resident), as well as considering fact-finding questions only, all with no difference between study conditions. The patient summary statement ('one-liner') was analyzed separately from the remaining questions, also with no difference noted between the two groups.

Given the importance of efficiency in clinical care, the time to completion for each scenario was recorded. Timing improved significantly between first and second attempts using both iNYP (p = 0.009) and HARVEST (p = 0.014), but there was no difference in improvement between the two conditions from attempt one to two (p = 0.959). There was also no

Figure 4: Screenshot of HARVEST for the same deidentified patient as in the iNYP screenshot in figure 1. For the selected time range (3 months), stable angina, pulmonary hypertension, end-stage renal disease, and dyspnea are the most prominently documented problems. HARVEST also identified diabetes mellitus, hypertension, and dyslipidemia as important problems. The problem 'dyspnea' was selected. The Notes panel lists all notes in the selected time range that mention this problem. Out of these, a cardiology consult note is selected and displayed in the lower right panel, with all mentions of dyspnea (and synonyms) highlighted. On the timeline, documentation of dyspnea is highlighted by purple bars, indicating that dyspnea was a particularly salient issue at that time as well as 6 months later.
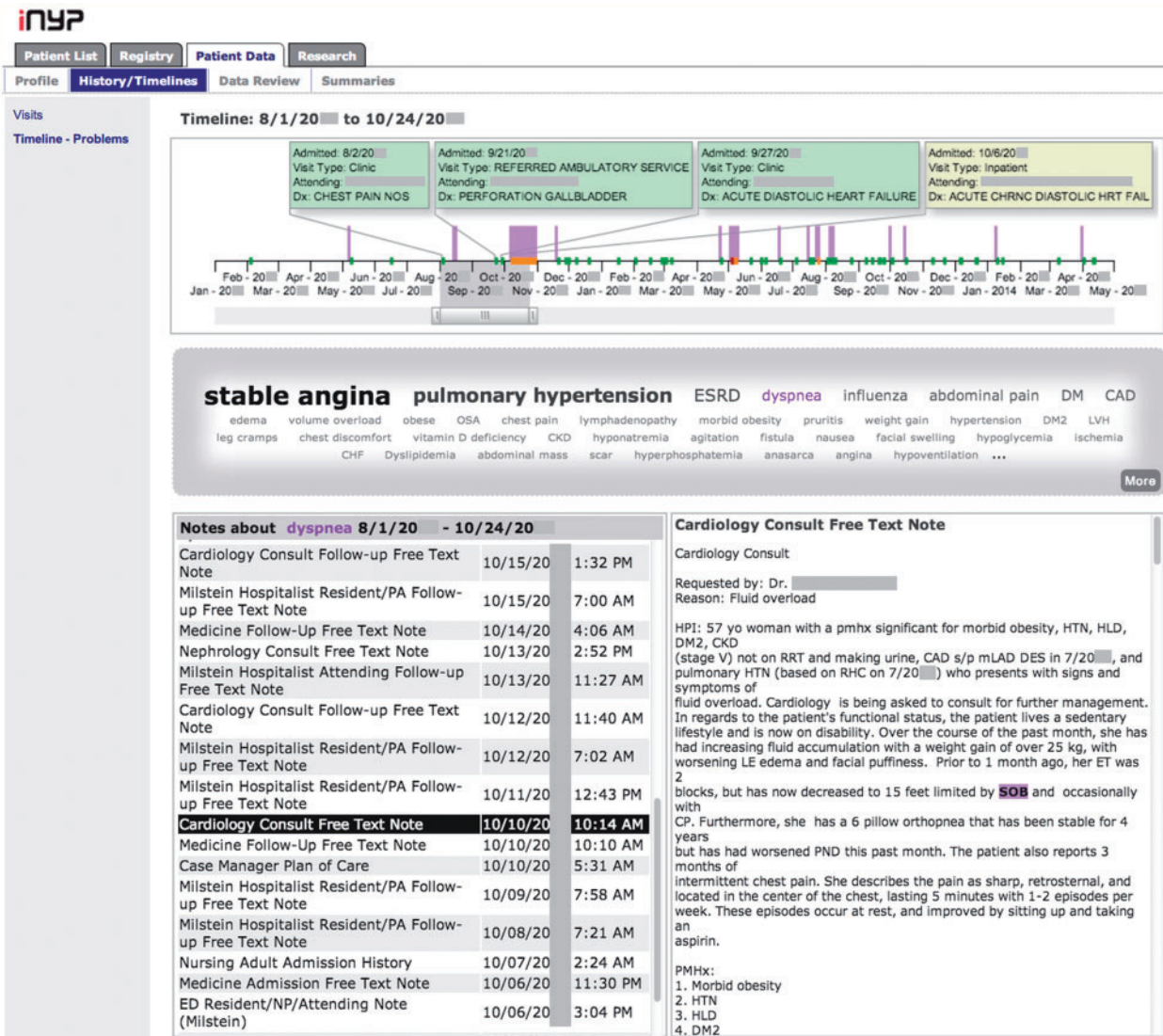
statistically significant difference in time to completion between iNYP and HARVEST conditions during the first attempt ($p = 0.151$), the second attempt ($p = 0.092$), and overall ($p = 0.333$), although the HARVEST condition tended to take longer (iNYP median/IQR 1015.0/239.0 s; HARVEST median/IQR 1098.0/102.6 s). Participants ran out of time in four iNYP- and six HARVEST-based scenarios ($p = 0.724$), although both groups had only a single incomplete questionnaire in the second trial (table 3).

Participant feedback

At the conclusion of the study, participants were asked a series of questions regarding their overall impression of HARVEST, including significant benefits or flaws in the tool.

Despite the novelty of the visualization, participants felt that it would be helpful in identifying subjective data not captured in objective, laboratory reports, and that HARVEST could help discover items that may otherwise be missed (table 4A). In particular, participants noted that the representation of a patient's

Table 2: Summary statistics for overall scores (normalized to 1) in the two testing scenarios

Condition	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
iNYP	0.57	0.67	0.72	0.74	0.80	0.91
HARVEST	0.35	0.70	0.72	0.71	0.75	0.84

Figure 5: Boxplot of overall questionnaire scores between the two scenarios (iNYP vs HARVEST). There were no statistically significant differences in scores between the two conditions.
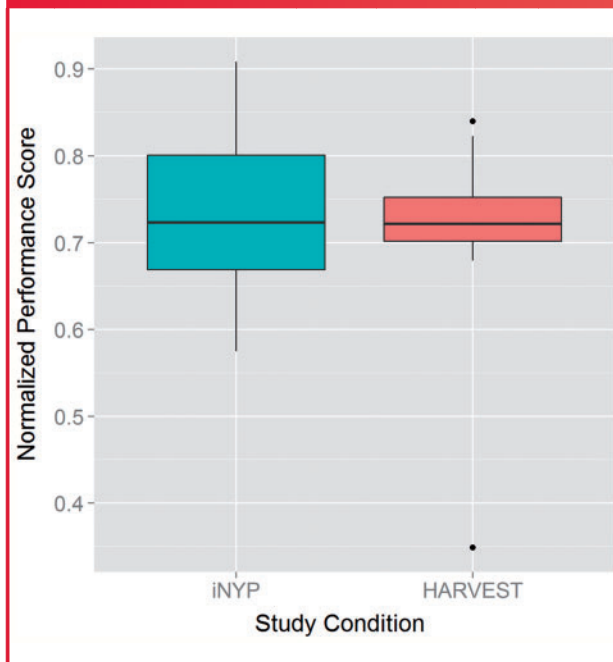
Table 3: Completion of task questionnaires in time allotted, by study condition and trial number

Condition	Task complete (number of trials)	Task incomplete (number of trials)
iNYP Overall	20	4
iNYP Trial 1	9	3
iNYP Trial 2	11	1
HARVEST Overall	18	6
HARVEST Trial 1	7	5
HARVEST Trial 2	11	1

longitudinal medical history is particularly useful when seeing a new patient in the clinic, admitting a patient to the hospital, or rapidly evaluating a patient in the ED (table 4B).

Three-quarters of participants said they would definitely use HARVEST again, while the remainder said they might access it. Using Likert-type questions, most participants found all three components of HARVEST somewhat or extremely helpful (83% for the timeline, 75% for the problem cloud, and 83% for note access) (figure 6). Those who found HARVEST components unhelpful commented on the lack of integration with other aspects of the health record, such as review of laboratory tests, as well as the inability to see all notes for a specific visit without selecting a problem (in the testing version of HARVEST, it was not possible to see all the notes for a given time period as only notes related to a selected problem appeared).

No correlation was present between overall satisfaction with HARVEST as evidenced by mean score on the three Likert-type post-task questions and overall mean score on all four patient cases ($r = 0.316$, $p = 0.318$) or mean score on patient cases with HARVEST ($r = 0.379$, $p = 0.224$).

Several consistent concerns regarding the tool emerged from study participants. The timeline in the HARVEST version used in the study extends back 2 years only, forcing participants to enter the iNYP interface to find older information. The extraction of problems from notes without regard to clinical relevance frequently populated the problem cloud. For example, negated terms (eg, 'no history of asthma') and early clinical thoughts (differential diagnoses) that do not ultimately pan out (eg, 'may have pneumonia') are stationed within the problem cloud alongside actual patient problems. While two-thirds of participants found no content redundant or unnecessary in HARVEST, the remainder remarked on the high number of problems related to review of systems in the cloud. Finally, each time a participant went from iNYP to HARVEST, the latter interface required reloading, causing a delay.

DISCUSSION

This paper reports on the development, implementation, and formative evaluation of a novel data summarization and visualization tool embedded directly on top of the EHR. With an appreciation of the growing complexity of EHR data across care settings and the increasing use of electronic documentation, HARVEST was designed as a point-of-care tool to support

Table 4: Subject feedback on the overall use of Harvest (A) and applicable usage in the clinical workflow (B)		
A.	▶	"It's a great adjunctive tool to visually represent the patient's chart history"
	▶	"Good visual representation of the patient's clinic and [ED] visits and admissions, and gives a good overall sense of the patient's medical problems"
	▶	"[Allows for] review [of] the medical record to find specific instances when things were diagnosed or managed"
	▶	"Useful tool to quickly tell burden of disease"
	▶	"Made me more confident that I wasn't missing information that can sometimes be buried in the list of [past medical history]"
	▶	"[I]t helped pick up on diagnoses within the chart that I otherwise would've had a lot of difficulty finding"
	▶	"The tool was very helpful in quickly getting a sense of how many (and what type of) encounters a patient had"
B.	▶	"[T]he Harvest tool would be most helpful when taking care of new patients and patient not already well-known"
	▶	"I would use it in pre-scrolling patients prior to seeing them both in the outpatient and inpatient setting"
	▶	"[Harvest] would allow me to better become acquainted with other people's patients in the event I was covering for them in clinic"
	▶	"[When] admitting a patient to the hospital, I feel like it would allow me to gather information to write a pertinent admission note in less time
	▶	"[I]n the emergency department this tool would allow me to get a rapid view of the important terms in the patient's medical record"

ED, emergency department.

Figure 6: Responses to Likert-type questions in the post-task questionnaire regarding helpfulness of the three features of HARVEST: access to the notes, the timeline, and the problem cloud. Most participants found the three features somewhat or very helpful (blue highlighted region to right of vertical bar), whereas fewer found the features somewhat or very unhelpful (pink and red highlighted region to left of vertical bar).
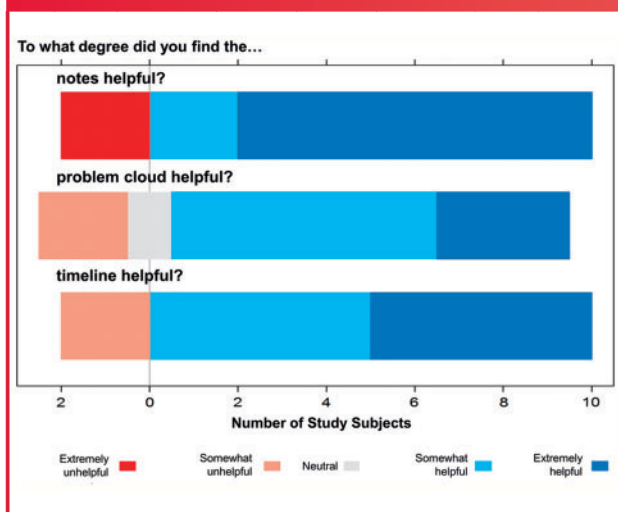
clinician workflow by leveraging the unstructured part of the record. As a summarization tool, HARVEST is not intended to replace other components of a medical record system, but rather to be used within and alongside the EHR. It may prove to be an extremely valuable adjunct to traditional clinical data review systems.

While there are established visualization principles to convey discrete and continuous numerical information, visualizing textual information in a useful fashion remains an open research question. The interactive time- and problem-oriented visualization of HARVEST, along with its easy and focused access to relevant documentation in the record is a contribution to this discipline. The use of scalable, distributed NLP for all patients in the institution and for a very wide range of problems is another contribution to informatics research. In contrast, traditional clinical NLP has been used for specific tasks and types of data to extract along with specific types of documents.[29,30]

New tools have the potential to disrupt ingrained workflows, and therefore have the potential to worsen performance initially. Using a task-based evaluation setup requiring examination of patient records, we found that physicians did equally well with HARVEST or iNYP, regardless of physician type, patient case, individual subject, or order of scenario. The same was observed for the recall task of the patient one-liner. The study subjects had never interacted with HARVEST before and were moreover power users of iNYP. Several subjects commented that, given time to interact with HARVEST, task performance and efficiency would likely improve. Despite its novelty and learning curve, which may have biased performance against HARVEST, physicians performed equally well with and without HARVEST without any statistically significant difference in efficiency (notably, while the task was timed with a maximum of 20 min, participants were not specifically instructed to attempt completion quickly). Future studies will

examine performance following real-world HARVEST use for several months.

The post-task questionnaire indicates high satisfaction with HARVEST. Physicians were enthusiastic about the tool and indicated they would be very likely to use it in their future daily activities. The tool made them more confident that they were not missing crucial information. They noted that HARVEST would be particularly useful when evaluating a new patient, as it provides a quick and reliable overview of a patient through time.

HARVEST design limitations and future work

Users provided feedback that fell into two broad categories: general tool functionality, which is amenable to relatively simple revisions, and purposeful design decisions, which have inherent tradeoffs.

As (nearly all) subjects noted, the ability to click on a visit in the timeline and automatically filter the problem cloud and notes within that particular visit was missing. Furthermore, in the case of complex patients, lack of a quick visit filter made the timeline sometimes too cluttered. Subjects mentioned that customization of the visualization either based on physician type or even individual physicians would be welcome. For instance, particular note types could be given a higher weight for computing individual problem's saliences (eg, for a nephrologist, problems mentioned in nephrology notes would be boosted). These functional criticisms are already being addressed in newer versions of HARVEST.

Some other criticisms, however, are more complex and rely on tradeoffs in information presentation and accessibility. Some participants remarked that the salience weights of problems rely heavily on their note frequency, without taking into account any a priori clinical importance of a problem. Some may have overly focused on prominent terms in the cloud, or in some cases on the contrary, ignored the prominence because they disagreed with frequency as an indication of importance. This criticism is taken into account. As a first proxy for salience, frequency is useful, but does not capture entirely clinical salience, and might even be misleading sometimes.

One participant raised the question that the timeline reflects the timing of documentation of problems, rather than a chronology of the problems themselves. Both types of timelines have their merits in a clinical support tool and it is an open question which timeline is preferable for patient record summarization. Recent advances in clinical NLP are promising to generate chronological timelines for longitudinal records,[31,32] but the time to produce them in a dynamic fashion is an important consideration for future work.

Currently, problems extracted from the notes are displayed independently of one another. Thus semantically related problems may be conveyed far away from each other in the cloud. We are exploring visualization techniques to either merge such problems or convey distance in the cloud to represent relatedness. Such functionality would not only convey a higher level of synthesis of the patient record, it might also address the feedback of some subjects about the overly large amount of

problems in some patient cases. How to gather the semantic relatedness of the problems in a robust, automated fashion is an open question of much interest to our research.

Finally, subjects commented on the lack of negation and uncertainty detection when extracting problems from the notes. It is still an open question whether these should be considered. In our experiments, we have not yet found any tools for negation detection that is robust enough for the large number of note types (and sublanguages) in our institution. More importantly, the fact that a clinician mentions a problem, even if negated, might be more informative than if not mentioned at all. Under this assumption, including negated problems is an acceptable summarization strategy.

Formative evaluation limitations

The task-based study had several limitations. First, this study compared a well-known interface against an entirely novel tool, without allowing significant time for familiarization. The study did not observe physicians during real-world workflow, although tried to mimic this setting with real patient cases, clinically relevant questions and tasks, and time pressure. Despite our prototyping the study, the first two questions proved ambiguous to some subjects, as 'the most prominent or important' conditions or complaints were asked. Some subjects understood importance along the acute versus chronic dimension, while others focused on the severity of disease diagnosis independent of its activity. Finally, because HARVEST is deployed into NYPH only thus far, our evaluation focused on patients from a single institution. As HARVEST was designed in an EHR agnostic fashion, relying only on an HL7 feed, porting to another EHR is feasible. We are eager to experiment with HARVEST in other institutions and assess reproducibility of our research.[33]

CONCLUSION

The amount of data in the EHR is growing rapidly, and innovations in data visualization and display have not kept up. Indeed, there is a real concern that information may be missed or lost, or that clinicians may have difficulty retrieving crucial data at the point of care. Whereas structured laboratory data can be more easily manipulated and displayed in tables, charts, and graphs, unstructured clinical documentation remains largely inaccessible and yet contains critical information. By facilitating access to a patient timeline, clinical problems extracted directly from notes, and the clinical documentation, HARVEST was designed to address the unmet need for clinicians at the point of care, facilitating review of essential patient information. In a formative study comparing the standard EHR with HARVEST, clinician performance and efficiency on clinically relevant tasks remain unchanged compared with the standard EHR, and users have provided considerable favorable and positive feedback on the tool. The deployment of HARVEST in our institution allows us to study patient record summarization as an informatics intervention in a real-world setting. It also provides an opportunity to learn how clinicians use the summarizer, enabling

informed interface and content iteration and optimization to improve patient care.

CONTRIBUTORS

JSH, CL, JST, and NE designed the evaluation study and analyzed the results. SLG, ES, DH, AE, MS, and NE designed and implemented HARVEST. SLG, DV, MS, and NE deployed HARVEST into the clinical information review system. JSH and NE wrote the manuscript.

FUNDING

COMPETING INTERESTS

None.

ETHICS APPROVAL

Columbia University Institutional Review Board.

PROVENANCE AND PEER REVIEW

Not commissioned; externally peer reviewed.

DATA SHARING

The study questionnaire is available to other researchers seeking additional information. A deidentified working example of HARVEST can also be accessed by interested researchers. This can be obtained by contacting the corresponding author.

REFERENCES

1. McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectability of man. *N Engl J Med*. 1976;295:1351–1355.
2. Christensen T, Grimsmo A. Instant availability of patient records, but diminished availability of patient information: a multi-method study of GP's use of electronic patient records. *BMC Med Inform Decis Mak*. 2008;8:1–8.
3. Chase HS, Kaufman DR, Johnson SB, *et al*. Voice capture of medical residents' clinical information needs during an inpatient rotation. *J Am Med Inform Assoc*. 2009;16: 387–394.
4. Singh H, Spitzmueller C, Petersen NJ, *et al*. Information overload and missed test results in electronic health record-based settings. *JAMA Intern Med*. 2013;173:702–704.
5. Shapiro JS, Kannry J, Lipton M, *et al*. Approaches to patient health information exchange and their impact on emergency medicine. *Ann Emerg Med*. 2006;48:426–432.
6. Adler-Milstein J, Bates DW, Jha AK. A survey of health information exchange organizations in the United States: implications for meaningful use. *Ann Intern Med*. 2011;154: 666–671.
7. Stead W, Lin H. *Computational technology for effective health care: immediate steps and strategic directions*. National Research Council of the National Academies, 2009.
8. Powsner SM, Tufte ER. Graphical summary of patient status. *Lancet*. 1994;344:386–389.
9. Feblowitz JC, Wright A, Singh H, *et al*. Summarization of clinical information: a conceptual model. *J Biomed Inform*. 2011;44:688–699.
10. Plaisant C, Mushlin R, Snyder A, *et al*. LifeLines: using visualization to enhance navigation and analysis of patient records. Proceedings of the Annual American Medical Informatics Association Fall Symposium (AMIA). 1998:76–80.
11. Shahar Y, Boaz D, Tahan G, *et al*. Interactive visualization and exploration of time-oriented clinical data using a distributed temporal-abstraction architecture. Proceedings of the Annual American Medical Informatics Association Fall Symposium (AMIA). 2003:1004.
12. Rogers JL, Haring OM. The impact of a computerized medical record summary system on incidence and length of hospitalization. *Med Care*. 1979;17:618–630.
13. Wilcox A, Jones SS, Dorr DA, *et al*. Use and impact of a computer-generated patient summary worksheet for primary care. Proceedings of the Annual American Medical Informatics Association Fall Symposium (AMIA). 2005: 824–828.
14. Flanagan ME, Patterson ES, Frankel RM, *et al*. Evaluation of a physician informatics tool to improve patient handoffs. *J Am Med Inform Assoc*. 2009;16:509–515.
15. Were MC, Shen C, Bwana M, *et al*. Creation and evaluation of EMR-based paper clinical summaries to support HIV-care in Uganda, Africa. *Int J Med Inf*. 2010;79:90–96.
16. Reichert D, Kaufman D, Bloxham B, *et al*. Cognitive analysis of the summarization of longitudinal patient records. Proceedings of the Annual American Medical Informatics Association Fall Symposium (AMIA). 2010.667–671.
17. Weed LL. Medical records that guide and teach. *N Engl J Med*. 1968;278:652–657 concl.
18. Weed LL, Weed L. *Medicine in denial*. Charleston, SC: CreateSpace Independent Publishing Platform, 2011.
19. Graedon T. Is Larry Weed right. *J Particip Med*. 2013;5:e13.
20. Cao N, Gotz D, Sun J, *et al*. SolarMap: multifaceted visual analytics for topic exploration. Proceedings of the IEEE 11th International Conference on Data Mining (ICDM). 2011: 101–110.
21. Blei DM, Lafferty JD. A correlated topic model of Science. *Ann Appl Stat*. 2007;1:17–35.
22. Hripcsak G, Cimino JJ, Sengupta S. WebCIS: large scale deployment of a Web-based clinical information system. Proceedings of the Annual American Medical Informatics Association Fall Symposium (AMIA). 1999:804–808.
23. Wilcox AB, Vawdrey DK, Chen Y-H, *et al*. The evolving use of a clinical data repository: facilitating data access within an electronic medical record. Proceedings of the Annual American Medical Informatics Association Fall Symposium (AMIA). 2009:701–705.
24. Li Y, Lipsky Gorman S, Elhadad N. Section classification in clinical notes using supervised hidden Markov model.

Proceedings of the 1st ACM International Health Informatics Symposium (IHI). 2010:744–750.

25. Lipsky-Gorman S, Elhadad N. *ClinNote and HealthTermFinder: a pipeline for processing clinical notes*. Columbia University Technical Report, 2011.

26. Bodenreider O, McCray AT. Exploring semantic groups through visual approaches. *J Biomed Inform*. 2003;36: 414–432.

27. The CORE Problem List Subset of SNOMED CT®. http://www.nlm.nih.gov/research/umls/Snomed/core_subset.html. Accessed May 1, 2014.

28. Manning CD, Raghavan P, Schutze H. *Introduction to information retrieval*. Cambridge: University Press, 2008.

29. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? *J Biomed Inform*. 2009;42:760–772.

30. Pradhan S, Elhadad N, South B, *et al*. Evaluating the state of the art in disorder recognition and normalization of the clinical narrative. *J Am Med Inform Assoc*. 2014. Published Online First: 21 Aug 2014. doi:10.1136/amiajnl-2013-002544

31. Styler W, Bethard S, Finan S, *et al*. Temporal annotation in the clinical domain. *Trans Assoc Comput Linguist*. 2014;2: 143–154.

32. Raghavan P, Fosler-Lussier E, Elhadad N, *et al*. Cross-narrative temporal ordering of medical events. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL). 2014:998–1008.

33. Kleinberg S, Elhadad N. Lessons learned in replicating data-driven experiments in multiple medical systems and patient populations. Proceedings of the Annual American Medical Informatics Association Fall Symposium (AMIA). 2013: 786–795.

AUTHOR AFFILIATIONS

[1]Department of Biomedical Informatics, Columbia University, New York, USA

[2]Division of Nephrology, Department of Medicine, Columbia University, New York, USA

[3]College of Physicians and Surgeons, Columbia University, New York, USA

[4]New York Presbyterian Hospital, New York, USA

RESEARCH AND APPLICATIONS