More Text Please! Understanding and Supporting the Use of Visualization for Clinical Text Overview

Nicole Sultanum, Michael Brudno
University of Toronto and Hospital for Sick Children
{nicolebs,brudno}@cs.toronto.edu

Daniel Wigdor
University of Toronto
daniel@dgp.toronto.edu

Fanny Chevalier
University of Toronto and Inria
fanny@dgp.toronto.edu

ABSTRACT
Clinical practice is heavily reliant on the use of unstructured text to document patient stories due to its expressive and flexible nature. However, a physician's capacity to recover information from text for clinical overview is severely affected when records get longer and time pressure increases. Data visualization strategies have been explored to aid in information retrieval by replacing text with graphical summaries, though often at the cost of omitting important text features. This causes physician mistrust and limits real-world adoption. This work presents our investigation into the role and use of text in clinical practice, and reports on efforts to assess the best of both worlds—text and visualization—to facilitate clinical overview. We report on insights garnered from a field study, and the lessons learned from an iterative design process and evaluation of a text-visualization prototype, MedStory, with 14 medical professionals. The results led to a number of grounded design recommendations to guide visualization design to support clinical text overview.

Author Keywords
Clinical text; medical narrative; text visualization; clinical overview; medicine; medical visualization; user study.

INTRODUCTION
Clinical practice is a deeply interpretive activity that seeks to apply scientific medical knowledge to the needs and circumstances of an individual patient [22]. A patient's occupation, daily habits, religion, and personal life history all contribute to determining how they will cope with illness and what treatment options may lead to a better quality of life [33]. *Clinical text*, which refers to text-based documentation stored in a patient's medical chart (*e.g.*, consultation notes, specialist assessments and hospital admission reports), plays a fundamental role in accommodating this contextual richness, providing the necessary flexibility and expressivity to effectively document and reason about a patient's illness trajectory [1, 12, 18, 26].

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

ISBN 978-1-4503-5620-6/18/04...$15.00

DOI: https://doi.org/10.1145/3173574.3173996

While templates are commonly used to help frame the clinical text [39] they vary wildly across physicians, and free text remains the common shared format [16]. Content present in clinical text encompasses most information needed to obtain a sufficient, actionable understanding of the patient's medical state (a process referred to as achieving *sufficient clinical overview* [3]), and is where physicians spend most of their time when reviewing a medical chart [41].

Due to the time-sensitive nature of medical practice, however, text usage presents a number of scalability challenges (*i.e.*, larger records become time-consuming to peruse) and computability challenges (*i.e.*, it is difficult to use unstructured text to automate tasks) that are detrimental to a physician's performance when trying to obtain an overview of a patient's medical history from their chart [7, 25]. As a result, summarization strategies have been proposed to facilitate access to medical charts and its clinical text, including natural language processing (NLP) and data visualization [36]. These strategies focus on extracting clinically relevant information from text and then presenting it in a simplified manner—be it a textual [14, 15] or graphical visualization summary [21, 37, 38].

While these efforts make important research contributions, we see a number of technical, practical, and conceptual shortcomings with the general directions being pursued. First, there is a trend towards employing increasingly complex summarization techniques that are potentially incomplete or error-prone. This raises issues of trust [25], introduces additional overhead due to the need for human verification by a data curator or the physicians themselves [21], and hinders the effective integration of such systems into actual clinical practice [36]. Second, information extraction has typically focused on only using "strictly medical" details. Thus, contextual information such as the current family situation, patient occupation, how the patient is coping with the disease, and hints as to how physicians perceive said patient, are often lost in the process [36].

To address these issues in a practical manner, we believe that a fundamentally different approach to clinical text presentation needs to be considered. Using data visualization to facilitate clinical overview from text is a valuable and effective approach, but it needs to better encompass the contextual richness present in clinical text and to uphold physician's trust in the visualizations provided. Thus, we argue that clinical text needs to be more visible and accessible, and visualizations more tightly integrated to it, in order to preserve contextual details of the original clinical text while optimizing its study.

We establish the foundations for this research direction via a user-centered design process, with our work and contributions including: (a) findings from a formative study with 8 physicians, encompassing in-situ observations and interviews, and grounded by relevant literature in medicine and HCI, which led to a preliminary list of design goals; (b) findings from an evaluation study of *MedStory*, a proof-of-concept prototype we developed to embody the emerging design goals, validated by 14 physicians; (c) a list of design recommendations for the use of visualizations to support clinical text overview, recollecting on lessons learned throughout the entire process.

RELATED WORK

Of most relevance to the present exploration is research characterizing the role and usage of clinical text in medical practice, as well as clinical data visualizations to support medical text.

Clinical Text as Medical Narratives

Clinical text content has been extensively studied under the term *medical narratives*, which refers to "qualitative and semi-quantitative data gathered by the physician" [44]. Medical narratives are regarded as a flexible, memorable, and expressive tool to holistically document patient encounters [22] and to facilitate temporal and analytical reasoning [12]. The flexible discourse of medical narratives is also able to encompass a rich array of semantic constructs, including complex temporal patterns [34] and nuanced uncertainty qualifiers for medical evidence (*e.g.*, differentiating subjective and objective observations with the use of active or passive voice) [18].

This flexibility, however, comes at a cost. Since medical narratives are primarily embodied as *free unstructured text*, the difficulty for a physician to study their content increases as medical records grow in size. The overall lack of standardization of such text also results in increased complexity for the automation of medical narratives [43], which have seen significant evolution but limited adoption in real practice [36].

Clinical Overview and Summarization

The process of achieving *clinical overview* has been informed by a number of field studies characterizing the activity. We also consider *clinical summarization*, which refer to tools and techniques to support or automate the clinical overview process towards an output, *i.e.*, a summary. A number of conceptual axes were proposed to describe the design space of clinical overview including *historical* vs. *here & now* (*i.e.*, referring to importance of temporal placement) and *comprehensive* vs. *minimal overview* (*i.e.*, the time-information balance to achieve sufficient overview) [3]. A conceptual model of clinical summarization was also proposed that includes five steps—*aggregation* of information sources, *organization* (*e.g.*, grouping and sorting), *reduction and transformation* (*i.e.*, culling or modifying data for simplified understanding), *interpretation* (requiring clinical knowledge), and *synthesis* (*i.e.*, interpretation that leads to insight and decision-making)—all of which can be automated to some extent [9]. Regarding the use of clinical text for overview and summarization, reading practices around paper-based medical records were also characterized into four reading scenarios (*first time reading, re-reading, searching for facts*, and *problem solving*) and three reading modes (*reading*, *skipping*, and *searching*), revealing that clinical text is versatile, useful, multi-layered, and allows for various manipulations [32].

These studies highlight the intricacies involved in the clinical overview process, and reflects on the significant challenges and limitations around summarization. These challenges include a *lack of trust* in automated results [13, 25, 27, 36], *insufficient flexibility for customization* [25], and *limited reasoning over temporal events* [36]. The aforementioned findings are supported by our own qualitative investigations, and are revisited in the discussion of our formative study findings.

Visualization Systems For Clinical Text Summarization

Text visualization for clinical text summarization features a long history of research explorations [36, 42], exhibiting an evolutionary trend towards information extraction features to structure free text. Early systems, such as Powsner and Tufte's graphical summary [40] and Lifelines [37, 38], relied on manually extracted databases of clinical text features (*e.g.*, lists of medical problems and dates when each occurred). These systems were followed by others that featured NLP automation, first with semi-automated and partially curated pipelines, such as in Timeline [6], the CLEF chronicle viewer [14, 15], and various works by Hsu *et al.* [2, 20, 21], and more recently, with fully automated NLP pipelines such as in HARVEST [17] and in Forbes *et al.*'s nurse shift summarization tool [10]. While considerable progress has been achieved in the information extraction front, this success has not translated into real world clinical usage, arguably due to the lack of golden standards necessary to foster physician confidence in their results [36].

Regarding the text visualization design itself, there is a less pronounced evolution. Many of the visual constructs introduced by early systems were revisited in later systems, such as: the *presence of a timeline for temporal overview* that is event- [2, 14, 15, 17, 20, 21], document- [6] or source-oriented [40]) which indicates the importance of temporal awareness; *faceting*, *i.e.*, grouping related information by event type [2, 14, 15, 20, 21, 37, 38, 40], synonyms [17], or source [6]; and *chronicle summaries* (*i.e.*, time-ordered events presented as succinct text) [14, 15, 40]. These features require structured data derived from well-defined semantic and clinical specifications (*e.g.*, a date and medical taxonomy code associated to a symptom mentioned in text to enable plotting an event on a timeline). This reliance on structured data is problematic, as clinical text encompasses non-explicitness [34]. A later visual concept, V-Model [34, 35], was proposed to tackle this contextual complexity via a dynamic, text-driven timeline, but has not yet been integrated into any real working system.

In this work, we seek to preserve the qualitative aspects of clinical text and to support physician trust of automated visual summaries by rooting visualization *around* clinical text in an interactive visual system to support clinical overview.

THE ROLE OF TEXT FOR CLINICAL OVERVIEW

Much research has focused on understanding different aspects of the clinical workflow and on the design of visualizations for clinical summarization. Little, however, has focused on providing a cohesive, holistic understanding of clinical text for clinical summarization visualizations, including a better understanding of the role of text in real practice and how to better design visualizations *around* it. To address this gap, we conducted a formative qualitative study to understand the medical practices surrounding clinical text and contextualize our findings with respect to relevant related work.

Our study involved 8 physicians, recruited on a volunteer basis from a range of clinical scenarios, including four different health institutions in Canada (spanning outpatient hospital clinics and private clinics), five areas of specialization (*i.e.*, 2 general practitioners (GPs), 3 genetic pediatricians, 1 developmental pediatrician, 1 gastroenterologist and 1 orthopedic surgeon) and spanning 2 to 30 years of medical practice after residence (12 years in average). We conducted semi-structured interviews with all participants, and a full-day observation of the clinical practice for five of them, amounting to a total of 25 hours (average of 4 hours each). We sought to answer the following research questions:

1. *Clinical workflow:* What is the role of text in clinical practice, and what activities does it facilitate?
2. *Structure:* What are the embodiment(s) and format(s) of clinical text?
3. *Tasks:* How do physicians use clinical text?
4. *Challenges:* What obstacles do physicians face with clinical text?

In what follows, we present a summary and contextualization of our findings and derive design goals.

The Clinical Workflow

The clinical workflow can be represented as a three-stage process consisting of *preparation*, *consultation*, and *wrap-up*. Preparation refers to the period spent studying records, before a consultation. Consultation refers to the period that the physician spends interacting with the patient. Wrap-up refers to the period spent consolidating the documentation after seeing the patient. Clinical text is used in each of these stages, although information needs differ slightly per stage.

In the preparation stage, physicians sought to obtain an understanding of the medical situation of the patient they are going to see. The specific information that each physician used to achieve overview varied by specialty, but preparation intervals were quite short overall, *i.e.*, between 2 to 10 minutes in primary and secondary care; in highly specialized care (*e.g.*, pediatric genetics) we observed 30-minute intervals, but these were reported to be atypical. As a whole, physicians explained that they are under constant time pressure, and tend to prioritize recent patient information to quickly situate themselves and establish a continuity of care. Most of the information sought out is stored in text format, including a physician's own summary notes, referral reports (letters sent by a GP to a specialist), discharge summaries, and specialist assessments.

During consultation and wrap up, we observed physicians document relevant aspects of the patient visit. This activity was sometimes supported by a directed search for clinical information, often in text documents, to seek specific details (*e.g.*, lab results), provide support and context for the information provided by the patient (*e.g.*, checking past episodes of chest pain given a recent episode), or establish correlations between current and past episodes (*e.g.*, reporting a reduction in cholesterol as a result of exercise). Locating information in text is time consuming, and we observed physicians disengaging from the patients to look up information in the records on several occasions.

Embodiments

In reviewing medical records (two institutions allowed us access to sample records) and observing physicians do their pre-consultation reviews, we found clinical text to be the primary source of information for physicians, existing in different formats and for different purposes. In some instances, there were thorough clinical reports, written to summarize a specialist's assessment or a patient's hospital stay in detail. In other instances, we observed the use of personal physician notes, handwritten with shorthand language, that were used as temporary documentation to be later used for dictation, or to support the creation of an official letter. We also observed other documents in the record such as daily inpatient reports (*i.e.*, concise assessments made during a hospital stay) with short-term usefulness to local healthcare staff, or a GP's referral letter to a specialist that was used to introduce the patient and to request for a certain condition to be investigated. Apart from a few domain-wide formats (*e.g.*, SOAP–Subjective, Objective, Assessment and Plan–which most medical assessments adhered to, if only roughly), document structure and depth varied significantly across institutions, specialties, and individuals.

While paper had a significant presence for shorthand summary notes or as hard copies of electronic documents, all visited institutions used electronic health records (EHRs). While no two institutions used the same EHR system, having records be represented as an episodic document list that was sorted by date or source was a common pattern. In terms of presentation, we also found some visualization support for structured data (such as lab measurements plotted in a graph) but little to no support for text data. In addition, the few summarization views that were present were manually maintained by physicians (*e.g.* a short list of medical problems for a patient).

Using Clinical Text

We found physicians studying clinical text in three main situations: while obtaining a general overview when seeing a new patient, while reacquainting oneself with a returning patient's record to activate memory triggers, or while answering patient-related questions during consultation and wrap-up.

Given the time pressure, physicians adopted filtering strategies to optimize their study. In our formative study, the most common strategy was to start from the most recent documents (when they exist), since they are the most likely place to find current, pressing issues. Documents were then quickly skimmed or more thoroughly read (depending on the complexity of the case and how familiar the physician was with the patient and the chart), with the occasional jump between documents that provided additional details. These observations align with previous findings on clinical reading practices [32].

From our observations and interviews, we found a few general, recurrent physician-driven questions, for which answers were often sought in clinical text:

- **Establishing focus:** What are the current complaints and for what reason is this patient seeking medical care today?
- **Gathering context:** What is the relevant medical history for the problem(s) at hand?
- **Getting the gist:** What are the salient medical and psychosocial problems for this patient that I should know?
- **Establishing continuity:** What happened during the last visit, and what requires follow-up?
- **Filling in the gaps:** What happened since the last visit, *e.g.*, investigations or hospitalizations?

These questions underline the importance of temporal awareness and temporal reasoning, with a prioritization of recent information for continuity of care [5, 41] and the need for a physician to piece together parts of a record that pertain to related issues (*e.g.*, find all text related to diabetes) [29]. These aspects characterize a *narrative construction* activity [22], encompassing low level temporal reasoning inquiries such as what events took place (narrative plot), when events took place and how they evolved (temporal awareness, trends), why they took place (causality), and how the events are related (correlations, relationships).

Challenges

The one pervasive challenge physicians faced lies in their *time constraints*, which limits the depth of overview physicians can reach. In addition, because most of the go-to information was stored in text documents, there is an additional reading overhead. Our observations agree with previous research on physician summarization tasks, which found that half of a physician's time was spent on clinical notes alone [41].

Another significant aspect of text usage is record *fragmentation*, *i.e.*, the episodic nature of medical documents that makes it difficult to piece illness trajectories together. This issue is also recurrently pointed out in the literature as an obstacle to *temporal awareness* [25, 29, 30] and *faceting* of the illness trajectory into sub-stories (*i.e.*, sets of events related to one particular condition) [3, 9].

The *flexibility* in documentation standards also arose as another significant challenge facing physicians. Despite its value in expressivity for the physician, flexibility also introduces complexity due to the wide variety of formats possible [16], making it more difficult for a physician to locate information in the text, or determine if the information is available [45].

Lastly, we noted that some sensitive contextual information related to medical care (*e.g.*, whether patients were known for being particularly difficult, non-compliant, or if there were any concerning family issues) was not directly recorded in the patient chart, although facts were widely known and shared verbally within the healthcare team. Similar behaviours were noted in observational studies on inpatient wards within healthcare teams [31, 48]. Upon closer inspection, we noticed that subtle hints on these factors regarding the patient's state of mind and behaviour (*e.g.*, if a patient is very non-compliant) could sometimes be inferred from more descriptive notes, but would hardly be highlighted, and would be difficult to find when only skimming the text. This is something we found worth investigating in the context of clinical text overview.

Design Goals

From the related literature and our formative study, we identified a number of recurrent issues, including *fragmentation*, *lack of trust*, *document structure trade-offs*, and the *prevalence of narrative* and *temporal awareness*. We distill this knowledge into five preliminary design goals:

(TX) ***Design for and around text*** by accommodating the unstructured nature of the original text to leverage its contextual richness.

(TM) ***Support temporal awareness*** by conveying how medical problems, treatments, and patient attitudes progressed over time.

(GR) ***Support levels of granularity*** to enable overview-to-details navigation and facilitate information retrieval for quicker insights.

(FC) ***Support multiple facets*** to mitigate record fragmentation by aggregating related topics for easier understanding and smoother narrative construction.

(TR) ***Foster overview trust*** by allowing efficient triangulation via extensive linking between text and corresponding visual abstractions.

While these factors have been discussed in isolation by past works, we believe this to be the first attempt to holistically consider such principles in the context of clinical overview support. Also, to our knowledge, no tool encompasses all of these design goals simultaneously, a gap we seek to fulfill.

MEDSTORY

To validate the proposed design goals, we created *MedStory*, as a text visualization proof-of-concept to be evaluated in a realistic clinical scenario. In our design explorations, we sought to emphasize more qualitative aspects of text, as well as ways to integrate natural language processing and visualization to support insight.

Data and Pre-processing

MedStory was implemented as a D3/React Web application with a Python server. To populate the system, we selected document sets for two different patients, each containing 5 *clinical notes*, taken from the i2b2 NLP 2014 Challenge dataset [23]. Clinical notes were pre-processed for relevant information retrieval, using a mixture of curated and automated methods. Manually curated data included (a) *section outlines* (*e.g.*, parts of the text under "Past Medical History" and "Medications"), (b) a *medical topics code list* (such as "cardiovascular" or "diabetes" issues) encompassing high-level parent taxonomy codes from SNOMED-CT [24] for topical classification and (c) a *high-level text summary* consisting of bullet point lists about recent medication and family/social history issues. Automated processing included (a) *sentence tokenization* and *sentiment analysis* [28] at a sentence level; (b) *noun phrase extraction* and subsequent association to *medical taxonomy codes* (SNOMED-CT) via *cTAKES* [11]; (c) *representative*

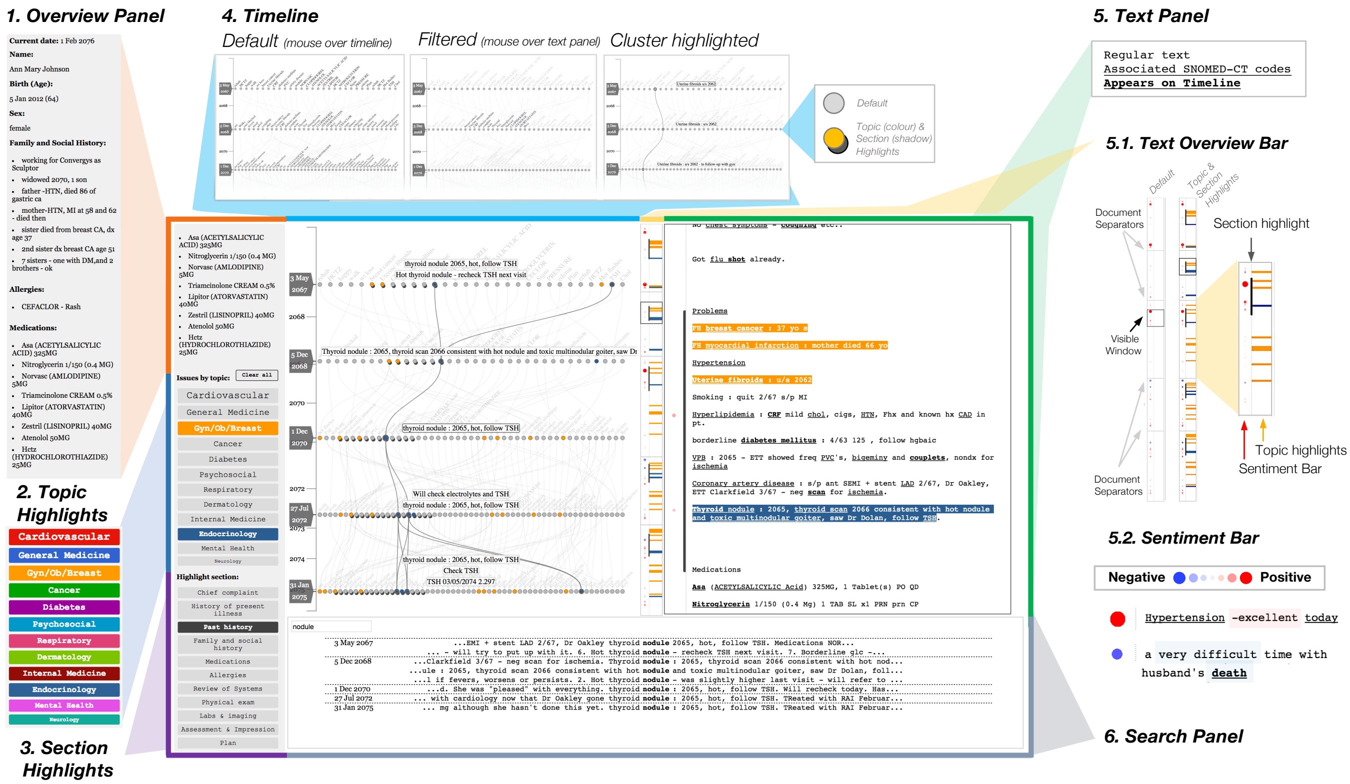


Figure 1: Medstory in action. Features include (1) the Overview Panel, (2) the Topic Highlights, (3) the Section Highlights, (4) the Timeline, (5) The Text Panel (containing (5.1) the Text Overview Bar and (5.2) the Sentiment Bar) and (6) the Search Panel.

noun phrase selection to serve as a short sentence-level summary, based on term frequency on a bag-of-words model computed from a set of 3598 clinical notes [23]; (d) *sentence classification* using the taxonomy codes in the sentences and the SNOMED-CT hierarchy to find which parent codes in the *medical topics code list* for each sentence; and (e) *sentence clustering* to identify recurring themes across clinical notes, using a greedy clustering algorithm that employed ontology codes and string similarity for grouping.

Interface Design

The interface of *MedStory* consists of four main components: (a) the **Overview Panel** (containing **Topic and Section Highlights**), (b) the **Text Panel** encompassing a **Text Overview Bar** and a **Sentiment Bar**, (c) the **Timeline**, and (d) the **Search Panel** (Figure 1). Panels were positioned so as to correspond to their levels of granularity (**GR**), with the Overview Panel providing high-level summarized information (left), the Timeline providing thematic snippets of text as medical events in chronological fashion (top middle), and the Text Panel providing the full content (top right). The Search panel is used for crosscut keyword search (bottom) (**FC**).

The first information point for the physician is the Overview Panel (Figure 1(1)), containing static summarized information about the patient, including demographics, family and social history, and a list of medications **(GR)**. The Topic Highlights list (Figure 1(2)) also provides a high-level overview of medical concerns that the physician may consider to inquire further (**GR**) Topics are sorted according to the frequency and length as they appear in the clinical notes (**TX**), which the physician can use as an indicator of prominence, since more significant issues tend to be more thoroughly documented.

From there, the physician may choose to narrow down information in the denser Timeline and Text Panel components via two types of filters, the Topic (Figure 1(2)) and Section Highlights (Figure 1(3)) (**FC**). Topic filters, when active, assign a specific colour to related sentences in the Text Panel and medical events in the Timeline (Figure 1(4)), whereas Sections are displayed as dark grey outlines (Text Overview Bar) and shadows (Timeline) (**TM, TR, FC**). Using the Topic Highlights, the physician can locate information related to medical subjects and problems (*e.g.*, related to a specialty or a particular disease, generated from the manually curated *medical topics code list*), whereas the Section Highlights provide a structural indicator for segments falling under common clinical text headings (*e.g.*, "Medications" or "Past Medical History"). When used concurrently, the physician can find answers for questions such as if anyone in a patient's family (Section) has a history of cancer (Topics) **(FC)**.

Following, the physician can probe for trends via the Timeline (Figure 1(4)), which provides a visualization of select text snippets and semantic interconnections as nodes and edges (with connections based on the *sentence clusters)* (**TM, FC**). Each horizontal track corresponds to a clinical note, with events anchored on a time axis and sorted by the order they appear in the text. By default, only a short label is shown for each event (*representative noun phrases*) for overview **(GR)** (Figure 1, *top right*) but the physician can reveal details and connections by hovering over event nodes (Figure 1, *bottom right*). To guide investigation, the physician can either hover over specific highlighted events (Topics and Sections), or traverse events in a note with the mouse. If curious about a certain event, the physician can click on an event node to be redirected to the corresponding sentence in the Text Panel, which gets highlighted in the text for easy spotting (**TX**, **TR**).

When the physician is ready to dive into details, they can move on to the Text Panel (Figure 1(5)), containing a contiguous list of all clinical notes (**TX**) and in chronological order for seamless navigation (**TM**). From the Text Overview Bar (Figure 1(5.1)), the physician can get a sense of where they are and where to find highlighted content (Topics as coloured horizontal highlights, and Sections as vertical dark outlines) (**FC**), which can be navigated to with a click. While in the Text Panel, the physician will also find qualitative text cues from the Sentiment Bar (Figure 1(5.2)), which convey sentence-level sentiment analysis as coloured bubbles beside corresponding sentences in the text (**TX**), indicating positive (red) and negative (blue) sentiment. These cues are meant to capture contextual hints about the patient that are sometimes present in the clinical text, as informed by our formative study. It was designed to be more suggestive than authoritative, with subtle glyphs that quickly fade away on smaller polarity values (both size and opacity are mapped to polarity magnitude). After looking at a few recent notes, the physician may be curious about when a certain issue was first mentioned, or if a medication has been discontinued. Bolded terms in the Text Panel are also present in the Timeline, and can be interacted with to reveal corresponding events in the Timeline **(TM)**.

Finally, the physician may look for specific terms (*e.g.*, items they suspect may have been missed by the Timeline) using the Search panel to perform keyword search across all clinical notes (Figure 1(6)) (**TR, FC**). Results are presented as a list of extended snippets (displaying a few words before and after the search result for context (**GR, TX**)), grouped by note and sorted chronologically by note date; this arrangement resembles a chronology, and from the preview content the physician may be able to find certain trends immediately (**GR, TM**). Snippets can also be clicked for further inspection, redirecting physicians to the original mention in the Text Panel (**TX, TR**).

USER EVALUATION

We conducted an evaluation of *MedStory* simulating real world conditions. Our goal was to qualitatively assess the proposed design principles that *MedStory* embodies rather than focus on particular widgets of the interface, and to better understand how these principles would fare in practice. In particular, we were interested in users' opinions of *MedStory's* features and how they handled mistakes in automated information retrieval.

Participants

General practitioners were recruited to participate in this study, given their broad medical interests, more qualitative focus, and more marked reliance on narrative text (as informed by our formative study). Fourteen participants (8 female) not involved

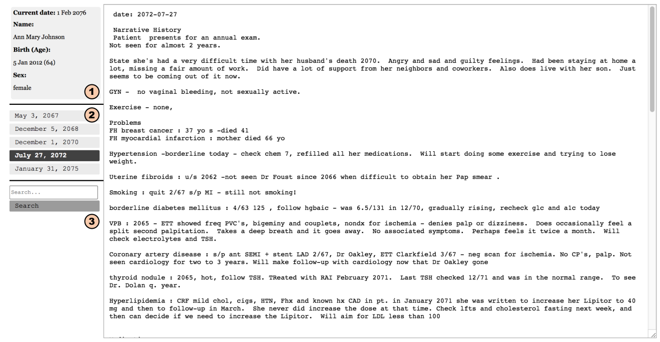


Figure 2: The baseline system, featuring (1) patient information panel, (2) a list of clinical notes, and (3) a search bar.

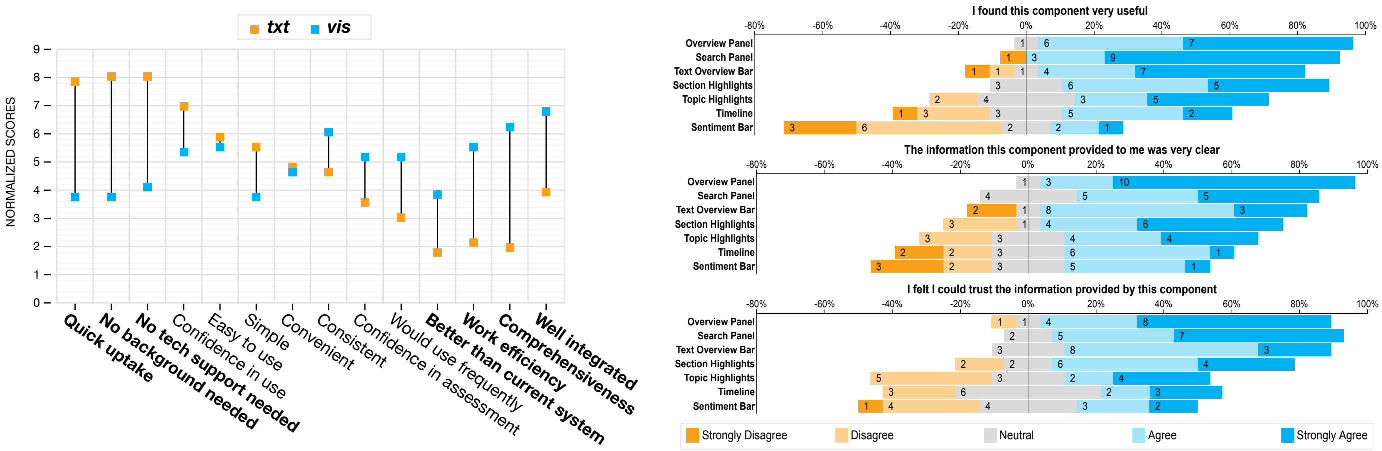
in our formative study were recruited via family medicine mailing lists, including 11 medical residents (7 in their 1st residency year, and 4 in their 2nd) and 3 family medicine physicians (each with 3, 10, and 11 years of experience), from 6 different health institutions. All participants reported using electronic health records regularly as part of their clinical work, and most estimated taking about 5 to 10 minutes to review a patient chart for a new patient. Participants received a $40 gift card for their participation.

Visualization Systems and Procedure

The study was designed to be both realistic and challenging, employing complex patient records and providing a short time for review, to elicit more ecologically valid feedback and to effectively bring forth pain points. In addition to using *MedStory*, we also included an additional session with a text-based system (Figure 2) to allow for a baseline comparison, *txt*, which was designed to resemble the standard medical record systems experience (as informed by our formative phase).

Study sessions comprised (a) a pre-study interview for participant profile and demographics, (b) a chart review session using the baseline system (*txt*), (c) a chart review session with *MedStory* (*vis*), and (d) a post-study questionnaire and semi-structured interview to revisit participants' experiences. Chart reviews comprised a (i) 6-minute study of a patient's clinical notes, (ii) a verbal summary of the findings, and (iii) an interactive question and answer (Q&A) segment with four questions that required the participant to use the system for information lookup, thus encompassing both *pre-consultation* and *in-consultation* information seeking scenarios. Two distinct document sets of similar size (5 text-heavy notes) and significant complexity were selected, alternated between conditions and participants for uniformity.

Chart review sessions were each preceded by a training session, to familiarize participants with the interfaces and study structure. Participants were also asked to think aloud, if they felt comfortable doing so without hindering their performance. From two pilot studies, we found participants also needed practice to accommodate to the chart review structure; we thus opted to fix condition order—*txt* then *vis*— for faster initial uptake, given the reduced complexity of the *txt* condition.



(a) (b)

Figure 3: (a) Normalized scores (higher is better) for the adapted System Usability Scale for the baseline condition (*txt*) and MedStory (*vis*) (**bolded** items: $p < .05$); and (b) Likert assessment of MedStory components.

Data Measures

The questionnaires consisted of an adapted version of the System Usability Scale (SUS) [4], used to compare user experiences with *vis* and *txt*, as well as a component-level Likert-scale assessment of the various features in *vis* to assess user perceptions in terms of *usefulness*, *clarity*, and *trust*. Individual SUS entries were normalized from a 1-5 Likert (or 5-1 reverse Likert) to a 0-10 scale (higher is better) and averaged per criteria across participants (Figure 3a). We also performed a Mann-Whitney test on raw Likert scores for the SUS dimensions and highlight those that yielded significant differences ($p < .05$) between *vis* and *txt* (Figure 3a).

The chart review sessions were audio and screen recorded. Audio data from interviews was transcribed, coded, and reorganized into representative themes using affinity diagramming, which informed the qualitative aspects of both conditions as well as the rationales for participant actions.

We also recorded a number of metrics for the Q&A segment, including *completion time*, *correctness* (based on whether all important points of an answer were addressed) and *thoroughness* of provided answers (based on demonstrated reasonable effort to build confidence in their answers, *e.g.*, by looking at several notes for *txt* and using multiple search mechanisms for *vis*). The *thoroughness* label did not apply to cases where participants strictly recalled from memory, even if their answer was correct. Dictations were also assessed in terms of *completeness*, by checking against a checklist specifically crafted for each document set (with the help of a domain expert who did not partake in the study) featuring a comprehensive list of noteworthy medical issues. The qualitative metrics (*correctness* and *thoroughness* for the Q&A plus *completeness* for the dictations) were evaluated by two independent coders and were used to assess information performance (access and retrieval) across both conditions.

Results

Four themes emerged from our analyses: (a) differences between the *txt* and *vis* conditions, (b) a component-level assessment of *MedStory*'s features, (c) reflections on automation trust and (d) on supporting qualitative richness.

For all subjective feedback (interviews), we provide the number of participants who expressed a given comment. While some of these counts may appear low, it is worth noting that they reflect only explicit remarks, not including those who agree with the topic but did not express an opinion; as such, they should be seen as lower bounds, but potentially higher.

TXT vs. VIS

We used a number of metrics to objectively compare the *txt* and *vis* conditions. Dictation had an average of 20.1 relevant medical issues identified by participants in the *vis* condition (std: 6.5) and 20.8 in the *txt* condition (std:4.9). As for Q&A performance, average time to answer was also roughly the same, with *56 seconds* on *txt* (std: *46 seconds*) and *59 seconds* on *vis* (std: *46 seconds*). In both cases, differences are small enough to consider them equivalent (given possible inter-participant variations at such small scale) but which could also be attributed to a lack of familiarity with *MedStory*, as we saw some participants struggle with some of the novel features.

The overall quality of the answers, however, was considerably higher in the *vis* condition, with *92.8%* (52/56) of answers marked *correct*, versus *75%* (42/56) for *txt*. Participants were generally more *thorough* in the *vis* condition as well, with *71.4%* (40/56) *thorough* investigations versus *58.9%* (33/56) in *txt*. We noted that in the *txt* condition, participants tended to give up investigating more challenging questions (*e.g.*, trend finding that required comparing against different notes) due to the limitations of the *txt* condition. Findings thus suggest promising *vis* performance in this area, enabling faster and more accurate answers. In terms of user satisfaction, the SUS scores appear to support this view as well, with the *vis* condition scoring higher on perceived efficiency, comprehensiveness, and confidence in the assessment (Figure 3a).

On the other hand, a few aspects were preferred in the *txt* condition. In the interviews, participants reported the *txt* condition to be simple (5/14), familiar (6/14), easy to learn (3/14) and easy to use (2/14). Compared to *vis*, some also found it less distracting (2/14), more compartmentalized (*i.e.*, well separated notes, as opposed to a continuum) (3/14) and appreciated the extra space reserved for text (2/14). Again, this feedback is also confirmed by the SUS scores, with the *txt* system scoring well on simplicity and learnability (Figure 3a, first 5 dimensions).

Overall, criticism to *vis* concerned aspects where *txt* fared better. For example, some participants recommended allocating more space to text (5/14): "*I like having the chart* (*i.e.*, Text Panel) *being the biggest part, because it's the most important part, the rest is just the navigation*" (P11). Others commented that the interface had many unfamiliar elements (4/14), which would require more time to get comfortable with than was allotted in the study session (15-20 minutes of training), *e.g.*: *"I'm a creature of habit, I think doctors are, so we like things a certain way, we get used to them"* (P10).

MedStory Components

While participants were walked through all features in *MedStory* during training, not every participant chose to use each of them in the chart reviews and Q&A. Among the non-static components (*i.e.*, excluding the Overview and Sentiment Bar), Section Highlights was the only component used by all 14 participants, followed by the TextBar (13/14), the Search Panel (12/14), and the Topic Highlights and Timeline (both 11/14). From the get-go, we see that not all components were equally favoured. The component Likert assessment (Figure 3b) informed how particular features were perceived in terms of usefulness, clarity, and trust: the three metrics are closely correlated in most cases, with the Overview and Search panels favoured, followed by the Topic and Section Highlights, and finally by the Timeline and Sentiment Bar.

High performers. Participants listed their preferred components as the Overview Panel (11/14), Search (9/14), and Text Overview Bar (9/14). Regarding the Overview Panel, a number of aspects may have contributed to its success. First, its immediacy: it provides hints of what to expect and look for in the record, "*I thought this [Overview] bar here was great, because you had very succinct information about the patient."* (P11)". Second, its familiarity: most participants (12/14) mentioned having a similar panel in their current practice, "*I liked the Overview Panel, it's something we are very used to in having"* (P2); however, in their practice, this panel is manually maintained, an effort justified as *"something that you are required to keep up to date for your own benefit"* (P10). It is interesting that few participants seemed to question the accuracy of the Overview Panel, which we affirm based on their interactions during the chart review sessions and their subjective feedback. Several participants suggested expanding the Overview Panel to also include past and current medical problems (7/14). The Search Panel was praised for aggregating results across all records (3/14), as well as for the contextual cues around the search results, which, in several cases, bypassed the need for direct note inspection (3/14). Despite its value, a few participants still suggested that the component take up less screen space, so as to favour the Text Panel instead.

Mixed Opinions. While a few participants liked the Topic (5/14) and Section (4/14) Highlights for Sections), reservations were also expressed for Topics (5/14), partly because it was not clear why certain elements were or were not highlighted (*e.g.*, respiratory issues tagged under "Cardiovascular"), what the intersectionality and semantic ambiguity between topics was (e*.g.,* the difference between *"general medicine and internal medicine"* (P2)), and why there was an overgenerality of certain topics, which failed to narrow content down (*e.g.*, "Cardiovascular"). The comparative assessment for Topics (Figure 3b) also shows a decrease in trust. One participant suggested being able to customize the topic tags and list, as *"every patient might have something a little bit different"* (P6). Another participant however, appreciated the information overlap, as *"good double taking of the information"* (P3). Sections was less contested, although some found it felt *"hit or miss"* (P2) when sections were missing as they could not determine if they were not present in the note or if the underlying algorithm failed.

Problematic. Finally, the Timeline and Sentiment Bar presented the most problems. In the case of the Timeline, issues were mostly design and usability related, with complaints about excessive clutter (9/14), small font size (3/14), and unclear criteria for event selection (2/14). Despite the issues, the temporal awareness it provided was found relevant: 9/14 participants highlighted positive elements of the Timeline, such as being able to see event connections and flowthrough (5/14) as well as the presence of a time scale (2/14). Suggestions for improvement included event filtering and grouping to reduce clutter (4/14), and reduced real estate to accommodate a larger text panel (5/14). Regarding the Sentiment Bar problems went beyond design issues. Many participants did not find value in it (7/14), while some reported forgetting to use it (2/14); we also did not observe the Sentiment Bar being used in the chart review sessions. Reported reasons encompassed perceived mislabelings (*e.g.*, it was unclear why a sentence was marked as positive), the bar not conveying a clear idea of the whole contextual picture, and the sentiment feedback not appearing significant to the particular cases used in the study.

On Automation and Trust

Other noteworthy observations included how participants reacted to natural language processing errors on the *vis* condition. In the interview, several participants stated frustration with the mistakes they found (6/14) or could not understand or trust the logic behind the automated choices (4/14). This eventually led to some abandoning the automated features and reverting to back to reliable features (*e.g.*, keyword search) (2/14) or spending time to verify correctness (1/14). On the other hand, many found that automation uncertainty interfered little with their assessment (5/14), with trust being built by double checking information in the notes (3/14), *e.g.*, *"I see the only hindrance was my testing the computer in how much I trusted it. And once I established, 'I can trust it on this', then it made it faster and made me feel like I was doing things more comprehensively"* (P12).

One important factor was how participants gauged their preconceptions and expectations, both in terms of the perceived limitations of technology as well as what they judged the role of automation to be. One participant had different levels of trust for topic highlights versus section highlights, with the latter seen as more "standard", easier to automate, and therefore more trustworthy: *"I don't trust this part, right here [Topic Highlights]. The section stuff I trust pretty much 100%, like 99% [...] because these words [the Section labels] are the words that most people use when they dictate these sections"* (P12). Some also placed less responsibility on the automation and more on the physician (2/14), *e.g.*, "*I'm reliant on it to just highlight things, but I think the interpretation of evidence is still up to us, right?*"(P2). One participant suggested the use of notifications with manual curation as a better interaction model with automation: *"... I have a list of medications that I maintain, and I can update. But if the machine recognizes a new medication ... it could give me a notification, like, 'this patient seems to have been started on this medication ... do you wanna update it in your system"'* (P12).

On Qualitative Richness, and Treating the 'Whole' Patient. During the interview, participants were also probed on whether they thought the *vis* condition favoured obtaining greater emotional awareness of the patient, and why. Feedback was mixed. Several participants did not see particular advantages (6/14), with some arguing that this awareness *"comes with just knowing the patient"* (P6) (3/14). One found the added complexity could strain the relationship (P11). All other participants (8/14), found it could be slightly helpful, by facilitating access to psychosocial and mental health issues via Topic highlighting (2/14), access to family and social history via Sections and Overview (4/14), or even the Sentiment Bar, given more time and relevant use cases (2/14). In addition, one participant mentioned that by providing easier and faster access to the information, more time could be dedicated to looking at the patient and building rapport.

Discussion

Overall, the study fulfilled our five preliminary design goals by confirming their relevance, and also helping to validate the overall concept that a text-centered visualization may be suitable to real world practice performance-wise. We also confirmed other aspects revealed in the formative study, such as the strong overall push to optimize interactions and minimize clicks (due to the time pressure in clinic), and unveiled others, including some resistance to the new paradigm due to unfamiliarity and the learning curve. In what follows, we take a deeper look at our preliminary design goals, and revisit them under the light of our study findings.

First, the study revealed the significance for text as a *grounding reference* (**TX**). This was particularly clear in view of the challenging scenario we created for participants, who were faced with a complex patient case under significant time pressure using an unfamiliar system: when expectations failed because new features did not work as expected, participants quickly reverted to "old habits". Past research in supporting novice users transition to expert users found similar behaviors [8]. Participants also recommended dedicating more space to the Text Panel in detriment of the other components, while they referred to the clinical note as "the most important part" of the system. This emphasizes the importance of supporting the clinical note as a central piece of the visualization.

Regarding levels of granularity (**GR**) and aggregation via faceting (**FC**), we found that both were appreciated as shortcuts to information, although the Search and Overview were of particular importance. The consistently positive feedback on the Overview Panel was initially surprising, given it is static and does not provide very substantial information to deserve much attention. While part of its success is due to familiarity, the reason why it was considered so important is the "immediate context" it enabled, helping set initial expectations and providing hints to guide further investigation. The Search component provided similar insight, by including a bit of context alongside search occurrences that bypassed the need to navigate to the note.

We can also reassess the shortcomings and criticism of the Timeline as a ramification of this immediacy standpoint (**TM, FC**): most of the information on the Timeline was not relevant to the problem at hand, and when it was, it was not easy to locate it. Participants pointed to having the Timeline take a more *active role* in inferring the physician's current information needs, for instance taking the current active filters, search keywords and Text Panel content to selectively expand or filter out events (**GR, FC**). Similarly, the Search Panel could also list related terms to complement keyword searches (e.g, synonyms or words that often appear together).

Feedback on the other two faceting components (**FC**), the Topic and Search Highlights, were mixed mostly due to trust concerns (**TR**) and dissatisfaction with the semantic breadth and intersectionality of certain topics (**GR**), *i.e.*, whether a mention of "shortness of breath" should be labeled under "Respiratory", "Cardiology" (as symptom of heart failure), or both. This touches on personal biases and expectations physicians had for organizing information, also affecting how they perceived automation performance, and suggests a need for customization tools. Allowing physicians to create their own topic filters would provide them with unprecedented power to optimize information uptake [47]. Data curation could also serve as a means to fix automation mistakes, train underlying automated components [46], and perhaps even contribute to increased trust in automation (**TR**). On the topic of trust (**TR**), past research argues that trust in automation is not an inherent quality of a system, but rather a *relationship to be fostered* [19], in which case we should look at ways to allow physicians to assess and adjust their own confidence in the automation. This needs to be easy to do, as any extra effort to triangulate information may be enough to dissuade automation uptake altogether. On the other hand, we found indications in the evaluation of *MedStory* that physicians would engage in extra efforts as long as the resulting benefits are clear and worthy [8], *e.g.*, the manual version of the Overview Panel physicians reported maintaining in their practice.

Regarding temporal awareness (**TM**), the time progression elements (*e.g.*, the flowthrough, the time axis) were considered useful despite the limitations of the current Timeline. Interestingly, the *situational aspect of temporal awareness,* or "where in time I am" in the record was noted as being useful. Among the participants who reflected on this, the compartmentalization of notes in the *txt* condition was stated to provide a better notion of "where I am" in the record, versus the *vis* condition. On one side, this indicates a positive immersiveness into the patient's illness story (**TX**), but also calls for extra visual cues to orient physicians (*e.g.*, more explicit connections between the Text Overview Bar and the Timeline's temporal axis).

Finally, while our attempt to leverage the more qualitative side of the clinical text via Sentiment was only partially embraced by participants, it nonetheless brought forth interesting perspectives. One is that while fostering empathy and good rapport with the patient are a physician's prerogative, simple tools can offer extra support to paint a clearer picture of the patient beyond the illness – such as the mental health or psychosocial topics and family or social history (**FC**). We also believe the Sentiment tool may be worth revisiting (**TX**) in a lower time-pressure condition. Given the current limitations of sentiment analysis to capture the more subtle qualitative hints present in clinical notes, a more fruitful direction may also be to indirectly assess mood and quality-of-life indicators from the set of medically-based information (*e.g.*, aggressiveness of symptoms and existence of chronic illnesses that indicate a risk of depression).

DESIGN RECOMMENDATIONS

In the previous section, we reiterated and reflected on our original design goals in light of the insights obtained in the evaluation study. In this section, we revisit this discussion while listing the important ramifications of the above discussion as design recommendations to create visualization systems and aids to support clinical text overview.

Make text a central piece of the visualization. Text is foundational to a physician's training and trade, fulfills expectations, and offers a "familiar place to return to". It also represents the "hard evidence", and the most trustworthy material available to serve as reference for patient care. Text and its affordances should be taken seriously and carefully considered when designing visualization systems for clinical text overview. Text should also be given significant screen real estate.

Provide immediate, in-a-glance content. In the context of the significant time pressure physicians are under when seeking clinical overview, "first impressions matter", and will often be the "only" impression they will be able to have. Providing static information upfront, is important to help set the stage and provides leads to guide initial search and exploration.

Mediate trust in automation. While natural language processing provides valuable tools to make text more tractable in large scale, its output should be amenable to scrutiny. Physicians should be able to *build* and *maintain* appropriate trust in automation, including gauging expectations on a personal level and in a continuous manner, gaining greater awareness of the inner workings of an automated component, and knowing how to work around its limitations. These affordances must be carefully designed not to lead to excessive overhead.

Minimize effort by making visualizations proactive. The designer should make use of the rich text semantic properties available in clinical text (*e.g.*, Topics and Sections) as well as the usage context (*e.g.*, what is the physician looking at right now) to preempt user information needs, minimize user clicks, and save the physician's time.

Support situational and temporal awareness. Temporal reasoning over clinical events is fundamental for clinical understanding and decision-making. As the temporal information is rooted in text, the designer should facilitate temporal awareness not only by revealing connections and trends present in the text, but also to provide awareness of "where" text lies in the greater temporal scale.

Enable low-effort curation and customization. Fine-grained information needs, personal preferences, and expectations for automated output can vary wildly across physicians. Therefore, it is important to allow for a certain level of customization, as well as the possibility to override automation mistakes. If curation is offered, this trade-off should be carefully assessed by the designer, and benefits to the physicians should be made palpable and clear.

CONCLUSION AND FUTURE WORK

Clinical text overview is a challenging problem, to which data visualization has significant room to contribute. In this work, we presented a broad and extensive investigation into this question, via an iterative design approach. Our activities included (a) an initial grounding from a formative phase, based on a multidisciplinary literature review and insights from field studies that led to a list of preliminary design goals; (b) the design, development and evaluation of *MedStory*, a tool we designed to help assess the design principles we derived, and (c) reflections on the design of text-centered visualizations for clinical overview, reframing our initial design goals as a list of design recommendations for visualization-based systems to support clinical overview.

There are also a number of relevant extensions and parallel investigations spanning from this work that have not been mentioned in the scope of our design reflections. First, extending our assessment to include *other medical specialties* would be an important step towards generality, as we have found some significant variations in work styles across formations. *Collaboration* is another relevant aspect to look into, as patient care traditionally involves coordination among several care providers, within and across specialties. From a design perspective, it would also be interesting to look at how this text-driven visualization approach can encompass the heterogeneity of the medical record to seamlessly incorporate data such as structured tables (*e.g.*, labs) and medical imaging to the core text narrative. Ultimately, we believe there is significant room for future investigations—to which this work is an important but only initial step— and we hope to see further exploration on the topic.

ACKNOWLEDGEMENTS

We would like to thank Brittney Johnstone, Alexander Lozano, Juliana de la Vega and Dr. Navindra Persaud for the valuable contributions on the medical side; Patricia Thaine for her contributions to early versions of this work; and Michelle Annet for manuscript feedback and editing support. We also thank the anonymous reviewers for their thoughtful comments and the DGP lab as a whole for the helpful discussions and ideas. This work was partly funded by a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant.

REFERENCES

1. Jørgen Bansler, Erling Havn, Troels Mønsted, Kjeld Schmidt, and Jesper Hastrup Svendsen. 2013. Physicians' progress notes. In *ECSCW 2013: Proceedings of the 13th European Conference on Computer Supported Cooperative Work, 21-25 September 2013, Paphos, Cyprus*. Springer, 123–142. DOI: http://dx.doi.org/10.1007/978-1-4471-5346-7_7

2. Vijayaraghavan Bashyam, William Hsu, Emily Watt, Alex AT Bui, Hooshang Kangarloo, and Ricky K Taira. 2009. Problem-centric Organization and Visualization of Patient Imaging and Clinical Data. *Radiographics* 29, 2 (2009), 331–343. DOI: http://dx.doi.org/10.1148/rg.292085098

3. Claus Bossen and Lotte Groth Jensen. 2014. How physicians 'achieve overview': a case-based study in a hospital ward. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 257–268. DOI: http://dx.doi.org/10.1145/2531602.2531620

4. John Brooke and others. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

5. PJ Brown, JL Marquard, B Amster, M Romoser, J Friderici, S Goff, and D Fisher. 2014. What do physicians read (and ignore) in electronic progress notes? *Applied clinical informatics* 5, 2 (2014), 430. DOI: http://dx.doi.org/10.4338/aci-2014-01-ra-0003

6. Alex AT Bui, Denise R Aberle, and Hooshang Kangarloo. 2007. TimeLine: visualizing integrated patient records. *IEEE Transactions on Information Technology in Biomedicine* 11, 4 (2007), 462–473. DOI: http://dx.doi.org/10.1109/titb.2006.884365

7. Tom Christensen and Anders Grimsmo. 2008. Instant availability of patient records, but diminished availability of patient information: a multi-method study of GP's use of electronic patient records. *BMC medical informatics and decision making* 8, 1 (2008), 12. DOI: http://dx.doi.org/10.1186/1472-6947-8-12

8. Andy Cockburn, Carl Gutwin, Joey Scarr, and Sylvain Malacria. 2015. Supporting novice to expert transitions in user interfaces. *ACM Computing Surveys (CSUR)* 47, 2 (2015), 31. DOI:http://dx.doi.org/10.1145/2659796

9. Joshua C Feblowitz, Adam Wright, Hardeep Singh, Lipika Samal, and Dean F Sittig. 2011. Summarization of clinical information: a conceptual model. *Journal of biomedical informatics* 44, 4 (2011), 688–699. DOI: http://dx.doi.org/10.1016/j.jbi.2011.03.008

10. Angus Forbes, Mihai Surdeanu, Peter Jansen, and Jane Carrington. 2013. Transmitting narrative: An interactive shift-summarization tool for improving nurse communication. In *Proceedings of the IEEE Workshop on Interactive Visual Text Analytics (TextVis)*.

11. The Apache Software Foundation. 2013. Apache cTAKES. (2013). http://ctakes.apache.org/

12. Trisha Greenhalgh and Brian Hurwitz. 1999. Narrative based medicine: Why study narrative? *Bmj* 318, 7175 (1999), 48–50. DOI: http://dx.doi.org/10.1136/bmj.318.7175.48

13. Trisha Greenhalgh, Katja Stramer, Tanja Bratan, Emma Byrne, Jill Russell, and Henry WW Potts. 2010. Adoption and non-adoption of a shared electronic summary record in England: a mixed-method case study. *Bmj* 340 (2010), c3111. DOI:http://dx.doi.org/10.1136/bmj.c3111

14. Catalina Hallett. 2008. Multi-modal presentation of medical histories. In *Proceedings of the 13th international conference on Intelligent user interfaces*. ACM, 80–89. DOI: http://dx.doi.org/10.1145/1378773.1378785

15. Catalina Hallett, Richard Power, and Donia Scott. 2006. Summarisation and visualisation of e-health data repositories. In *UK E-Science All-Hands Meeting, 18-21 Sept 2006, Nottingham, UK*.

16. Kristiina Häyrinen, Kaija Saranto, and Pirkko Nykänen. 2008. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *International journal of medical informatics* 77, 5 (2008), 291–304. DOI: http://dx.doi.org/10.1016/j.ijmedinf.2007.09.001

17. Jamie S Hirsch, Jessica S Tanenbaum, Sharon Lipsky Gorman, Connie Liu, Eric Schmitz, Dritan Hashorva, Artem Ervits, David Vawdrey, Marc Sturm, and Noémie Elhadad. 2014. HARVEST, a longitudinal patient record summarizer. *Journal of the American Medical Informatics Association* (2014), amiajnl–2014. DOI: http://dx.doi.org/10.1136/amiajnl-2014-002945

18. Pamela Hobbs. 2003. The use of evidentiality in physicians' progress notes. *Discourse Studies* 5, 4 (2003), 451–478. DOI: http://dx.doi.org/10.1177/14614456030054001

19. Robert R Hoffman, Matthew Johnson, Jeffrey M Bradshaw, and Al Underbrink. 2013. Trust in automation. *IEEE Intelligent Systems* 28, 1 (2013), 84–88. DOI: http://dx.doi.org/10.1109/mis.2013.24

20. William Hsu, Corey W Arnold, and Ricky K Taira. 2010. A neuro-oncology workstation for structuring, modeling, and visualizing patient records. In *Proceedings of the 1st ACM International Health Informatics Symposium*. ACM, 837–840. DOI: http://dx.doi.org/10.1145/1882992.1883120

21. William Hsu, Ricky K Taira, Suzie El-Saden, Hooshang Kangarloo, and Alex AT Bui. 2012. Context-based electronic health record: toward patient specific healthcare. *IEEE Transactions on information technology in biomedicine* 16, 2 (2012), 228–234. DOI: http://dx.doi.org/10.1109/titb.2012.2186149

22. Kathryn Montgomery Hunter. 1991. *Doctors' stories: The narrative structure of medical knowledge*. Princeton University Press.

23. i2b2 tranSMART Foundation. 2013. Informatics for Integrating Biology & the Bedside. (2013). https://www.i2b2.org/

24. SNOMED International. 2002. SNOMED-CT. (2002). http://www.snomed.org/snomed-ct

25. Lotte Groth Jensen and Claus Bossen. 2016. Factors affecting physicians' use of a dedicated overview interface in an electronic health record: The importance of standard information and standard documentation. *International journal of medical informatics* 87 (2016), 44–53. DOI: http://dx.doi.org/10.1016/j.ijmedinf.2015.12.009

26. Stephen B Johnson, Suzanne Bakken, Daniel Dine, Sookyung Hyun, Eneida Mendonça, Frances Morrison, Tiffani Bright, Tielman Van Vleck, Jesse Wrenn, and Peter Stetson. 2008. An electronic health record based on structured narrative. *Journal of the American Medical Informatics Association* 15, 1 (2008), 54–64. DOI: http://dx.doi.org/10.1197/jamia.m2131

27. Thai Le, B Reeder, H Thompson, G Demiris, and others. 2013. Health providers' perceptions of novel approaches to visualizing integrated health information. *Methods of information in medicine* 52, 3 (2013), 250–258. DOI: http://dx.doi.org/10.3414/me12-01-0073

28. Steven Loria. 2013. TextBlob: Simplified Text Processing. (2013). https://textblob.readthedocs.io

29. Troels Mønsted. 2015. Keeping Distributed Care Together: Medical Summaries Reconsidered. In *ECSCW 2015: Proceedings of the 14th European Conference on Computer Supported Cooperative Work, 19-23 September 2015, Oslo, Norway*. Springer, 143–161. DOI: http://dx.doi.org/10.1007/978-3-319-20499-4_8

30. Troels Mønsted, Madhu C Reddy, and Jørgen P Bansler. 2011. The use of narratives in medical work: a field study of physician-patient consultations. In *ECSCW 2011: Proceedings of the 12th European Conference on Computer Supported Cooperative Work, 24-28 September 2011, Aarhus Denmark*. Springer, 81–100. DOI: http://dx.doi.org/10.1007/978-0-85729-913-0_5

31. Glenn Munkvold and Monica Divitini. 2006. From storytelling to reporting-converted narratives. In *Proceedings of MCIS'06, the Mediterranean Conference on Information Systems*.

32. Else Nygren and Peter Henriksson. 1992. Reading the medical record. I. Analysis of physician's ways of reading the medical record. *Computer methods and programs in biomedicine* 39, 1 (1992), 1–12. DOI: http://dx.doi.org/10.1016/0169-2607(92)90053-a

33. Julian Oates, W Wayne Weston, and John Jordan. 2000. The impact of patient-centered care on outcomes. *Fam Pract* 49 (2000), 796–804.

34. Heekyong Park and Jinwook Choi. 2012. V-model: a new innovative model to chronologically visualize narrative clinical texts. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 453–462. DOI:http://dx.doi.org/10.1145/2207676.2207739

35. Heekyong Park and Jinwook Choi. 2014. V-Model: a new perspective for EHR-based phenotyping. *BMC medical informatics and decision making* 14, 1 (2014), 90. DOI: http://dx.doi.org/10.1186/1472-6947-14-90

36. Rimma Pivovarov and Noémie Elhadad. 2015. Automated methods for the summarization of electronic health records. *Journal of the American Medical Informatics Association* 22, 5 (2015), 938–947. DOI: http://dx.doi.org/10.1093/jamia/ocv032

37. Catherine Plaisant, Brett Milash, Anne Rose, Seth Widoff, and Ben Shneiderman. 1996. LifeLines: visualizing personal histories. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 221–227. DOI: http://dx.doi.org/10.1145/238386.238493

38. Catherine Plaisant, Rich Mushlin, Aaron Snyder, Jia Li, Daniel Heller, and Ben Shneiderman. 1998. LifeLines: using visualization to enhance navigation and analysis of patient records. In *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 76. DOI: http://dx.doi.org/10.1016/b978-155860915-0/50038-x

39. Stephanie E Pollard, Pamela M Neri, Allison R Wilcox, Lynn A Volk, Deborah H Williams, Gordon D Schiff, Harley Z Ramelson, and David W Bates. 2013. How physicians document outpatient visit notes in an electronic health record. *International journal of medical informatics* 82, 1 (2013), 39–46. DOI: http://dx.doi.org/10.1016/j.ijmedinf.2012.04.002

40. Seth M Powsner and Edward R Tufte. 1994. Graphical summary of patient status. *The Lancet* 344, 8919 (1994), 386–389. DOI: http://dx.doi.org/10.1016/s0140-6736(94)91406-0

41. Daniel Reichert, David Kaufman, Benjamin Bloxham, Herbert Chase, Noémie Elhadad, and others. 2010. Cognitive analysis of the summarization of longitudinal patient records. In *AMIA annu symp proc*, Vol. 2010. 667–671.

42. Alexander Rind, Taowei David Wang, Wolfgang Aigner, Silvia Miksch, Krist Wongsuphasawat, Catherine Plaisant, Ben Shneiderman, and others. 2013. Interactive information visualization to explore and query electronic health records. *Foundations and Trends® in Human–Computer Interaction* 5, 3 (2013), 207–298. DOI: http://dx.doi.org/10.1561/1100000039

43. S Trent Rosenbloom, Joshua C Denny, Hua Xu, Nancy Lorenzi, William W Stead, and Kevin B Johnson. 2011. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *Journal of the American Medical Informatics Association* 18, 2 (2011), 181–186. DOI: http://dx.doi.org/10.1136/jamia.2010.007237

44. Huibert J Tange, Arie Hasman, Pieter F de Vries Robbé, and Harry C Schouten. 1997. Medical narratives in electronic medical records. *International journal of medical informatics* 46, 1 (1997), 7–29. DOI: http://dx.doi.org/10.1016/s1386-5056(97)00048-8

45. Huibert J Tange, Harry C Schouten, Arnold DM Kester, and Arie Hasman. 1998. The granularity of medical narratives and its effect on the speed and completeness of information retrieval. *Journal of the American Medical Informatics Association* 5, 6 (1998), 571–582. DOI: http://dx.doi.org/10.1136/jamia.1998.0050571

46. Gaurav Trivedi, Phuong Pham, Wendy Chapman, Rebecca Hwa, Janyce Wiebe, and Harry Hochheiser. 2017. An Interactive Tool for Natural Language Processing on Clinical Text. *arXiv preprint arXiv:1707.01890* (2017).

47. Lauren Wilcox, Jie Lu, Jennifer Lai, Steven Feiner, and Desmond Jordan. 2010. Physician-driven management of patient progress notes in an intensive care unit. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1879–1888. DOI: http://dx.doi.org/10.1145/1753326.1753609

48. Xiaomu Zhou, Mark S Ackerman, and Kai Zheng. 2009. I just don't know why it's gone: maintaining informal information use in inpatient care. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2061–2070. DOI: http://dx.doi.org/10.1145/1518701.1519014