ChartWalk: Navigating large collections of text notes in electronic health records for clinical chart review

Nicole Sultanum, Farooq Naeem, Michael Brudno, and Fanny Chevalier

Abstract— Before seeing a patient for the first time, healthcare workers will typically conduct a comprehensive clinical chart review of the patient's electronic health record (EHR). Within the diverse documentation pieces included there, *text notes* are among the most important and thoroughly perused segments for this task; and yet they are among the least supported medium in terms of content navigation and overview. In this work, we delve deeper into the task of clinical chart review from a data visualization perspective and propose a hybrid *graphics+text* approach via *ChartWalk*, an interactive tool to support the review of text notes in EHRs. We report on our iterative design process grounded in input provided by a diverse range of healthcare professionals, with steps including: (a) initial requirements distilled from interviews and the literature, (b) an interim evaluation to validate design decisions, and (c) a task-based qualitative evaluation of our final design. We contribute lessons learned to better support the design of tools not only for clinical chart reviews but also other healthcare-related tasks around medical text analysis.

Index Terms—Electronic Health Record (EHR), Text Visualization, Close+Distant Reading, Clinical Overview, Medicine

1 Introduction

Healthcare workers spend increasingly more time and effort dealing with the EHR (electronic health record), which is currently leading to EHR burnout [2]. Reviewing existing documentation in the EHR accounts for approximately 32% of the total time spent in EHR [2]; the majority of this time is spent reading semi-structured *text* documentation [53], making it a time consuming and tedious process. The challenge is exacerbated by the variable quality of notes [3], and by limited content retrieval capabilities in most commercial EHRs [60]. This challenge is undertaken on a regular basis as healthcare workers perform a review of clinical notes when preparing to see a patient for the first time, a process called *chart review*—i.e., a deep dive into a single patient record to learn sufficiently about this patient's current and past medical history to guide healthcare decisions [60]. In view of this task and the associated burdens, a question arises on whether this task can be more readily supported via a visual analysis strategy.

And indeed, data visualization works have tackled the text overview problem in the context of clinical chart review with some success [49, 54, 64]. These works provide visual summaries of a patient's medical history and their health problems, with increased levels of scalability and sophistication via the use of natural language processing (NLP) to extract medically relevant details [49]. However, most of these works overemphasize the graphical overview aspects in the designs, with limited resources to trace visual elements back to the original text mentions they represent. This is detrimental to the chart review process, as clinical notes contain context that is key to obtain, and they possess unique affordances that are not as easily conveyed in visual form [60].

We argue that the chart review process calls for a more holistic and balanced visual analysis tool that elevates text to be a first class citizen, and that enables seamless navigation between visual summary items and their associated sources. In this work, we share our user-centered, iterative design process to create *ChartWalk*, a visual analysis tool that embodies this **graphics+text** concept to support in-depth chart reviews of very large and complex patient records, powered by medical natural language processing analysis. This work builds on past tools by the same authors that follow the same graphics+text philosophy — *MedStory* [60] and *Doccurate* [61] — extending them in *scope* (i.e., task coverage) and *scalability* (i.e., dataset size). From an initial set of design requirements informed by prior work (Section 4), we undertook two full iterative design cycles, each encompassing a design, development, and validation stage (Sections 5 and 6). We contribute findings gathered throughout the design process that collectively distill feedback from a diverse set of 19 healthcare practitioners, as well as consolidated findings and lessons learned from our final, in-depth task-based qualitative evaluation (Section 7).

- *Nicole Sultanum, Michael Brudno and Fanny Chevalier are with the University of Toronto. Emails: {nicolebs, brudno, fanny}@cs.toronto.edu*
- *Farooq Naeem is with the Centre for Addiction and Mental Health (CAMH). Email: farooq.naeem@camh.ca*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

2 Chart Review in Clinical Practice

Chart review has been broadly defined as the activity of reviewing "previously recorded data to answer clinical queries" [56]. While the term is often used to refer to cohort-based, large scale retrospective studies aimed at answering medical research questions [63], in this work we specifically refer to the in-depth study of a single patient's medical record for the purposes of care. In this context, the usual goal is to achieve sufficient overview [7], i.e., an "actionable understanding of the patient's medical state" to guide healthcare decisions [60].

The chart review consists of "collecting, distilling, and synthesizing" information from various parts of the EHR [16] the majority of which lies in semi-structured medical text reports [53]. In small scale, text makes for a rich, compact, flexible medium that seamlessly mediates context, nuance, and temporal reasoning [26, 30, 55]. As numbers increase, perusal becomes significantly more challenging, which is aggravated by a *lack of standardization* between individual document pieces [55]. The longitudinal and distributed nature of medical documentation ultimately leads to a heavily *fragmented and redundant* picture of the patient's illness trajectory that is challenging to piece together [40]. This is further exacerbated by the *occasional conflicting, erroneous, or redundant, or missing information* present in records [3], and the *stringent timeframes* around healthcare work [41]. Past assessments on chart review practices listed ways to tackle these challenges, including an emphasis on recent notes [53], a need to effectively aggregate related information from various sources [7] and around particular findings of importance [53, 60], and support seamless transitions between overview-level findings to detailed information [7, 60].

The chart review process and related tasks have been described by a number of qualitative studies. Nygren and Henriksson [45] looked at how physicians perused paper medical records, finding several unique reading scenarios still applicable to electronic records: *First time reading*, *re-reading*, *searching for facts*, and *problem solving*. This showcases the versatility of text to support a variety of information seeking patterns. Feblowitz et al. [17] proposed a conceptual model of clinical overview via five steps: *organization* (e.g., grouping and sorting), *reduction and transformation* (i.e., culling or modifying data for simplified understanding), *interpretation* (i.e., requiring clinical knowledge), and *synthesis* (i.e., interpretation that leads to insight and decision-making). These two works characterize chart review as a multifaceted sensemaking activity that can benefit from a diverse set of data operations [42].

3 RELATED WORK

We review works in data visualization that aim to address the chart review process and associated data and process challenges, followed by relevant graphics+text works in other analogous text-centered domains.

3.1 Data Visualization and Clinical Notes

In view of the sensemaking challenges associated with the chart review process, there is a long history of data visualization and content summarization strategies for clinical overview [49, 54, 64]. Works like Powsner and Tufte's Graphical Summary [52], Lifelines [50, 51], Dabek et al.'s Timeline [12], PatientExploreR [20], and V-Model [48] provided high-level graphical representations for overview of patient histories; however, they either relied solely on EHR structured data or on human curated data to isolate meaningful data features, while providing limited navigation to the original notes.

In parallel, natural language processing had been progressively introduced into graphical overviews to help scale the extraction of relevant overview details and populate timelines, from extracting medical problems [10, 25] to detecting medical events [22, 23, 28]. On the other hand, the use of "unvetted" automated approaches in clinical practice raises issues of trust, which in turn cannot be fully vetted without in-practice use [49]. As such, triangulation approaches to automation that support verification against original document sources may offer transparency over automated outcomes and help foster trust and needed fall-backs when dealing with products of automation, something our past works [60, 61] and this one sought to address.

We also emphasize the role of provenance in providing seamless navigation back to the original notes, as they often retain important context that is hidden from summary views [60]. Notable works that explored this **graphics+text** mix have provided design inspiration to our work. They include MedStory [60], our past work, which provided a small-scale proof of concept for a multi-faceted navigation of a small patient record (up to 5 notes); we extend this work by significantly increasing the supported scale and complexity of patient records. Among those that supported larger patient records, there is HARVEST [25], a tool that augmented a list of medical notes with a wordcloud of extracted medical issues and a timeline of events, and Doccurate [61], a past work of ours that provided fine-grained faceting features via user-defined semantic search filters. The present work extends them by providing more granular overview panels, including more detailed overview of temporal patterns and more fine-grained content organization approaches.

3.2 Data Visualization and Other Temporal Text Collections

Outside of the medical domain, there is a wider range of works in text visualization [36] and text visual analytics [38]. Specifically among visual analysis of *time*-indexed text collections (i.e., more akin to patient records), past works have looked at providing visual summaries and exploratory search for microblogging [13–15, 67], emails [39], product reviews [29], news [24], and academic literature [4, 37]. These tools aim to provide a high level overview of trends in text but are less suited or not designed for low-level text analysis that requires extensive reading of individual documents.

Conversely, text analysis works for the social sciences have leaned more towards a mixed **graphics+text** strategy, approached as a way to bridge both close and distant reading of very extensive single documents (e.g., books) [32, 33]. Notable examples include TextViewer [11], VarifocalReader [35], and Serendip [1]; all explicitly designed to support multiple levels of detail with clear mappings between overview items and the original text. Other works in investigative analysis [31, 57, 62] and investigative journalism [9, 59] also account for the revision of individual documents, albeit with an emphasis on pattern finding and user-defined structures. Visualizations of text conversations [19, 44, 66] provide text snippets in context of larger content threads. A shared aspect among these works and the healthcare domain is the need for context, causality and nuance present in text, which are difficult to convey via isolated and de-contextualized terms (i.e., wordclouds) or visual glyphs. While there is otherwise limited transferability between these domains and the medical domain, they provide design inspiration and help consolidate our mixed **graphics+text** approach further.

4 ITERATIVE DESIGN PROCESS: SETTING THE STAGE

In past works, we established a progression of increased understanding and support to the task of chart review. Beginning with MedStory [60], we created a proof of concept tool for a small record to validate design principles for a **graphics+text** oriented approach to chart review. This work informed a list of design requirements which were validated in a comparative study (and which form the basis of the requirements presented in Section 4.2) but the prototype had limited scalability and could display a maximum of only 5 notes. With Doccurate [61], we successfully scaled up in data size (from 5 to 300+ documents) and offered user-friendly tools to curate semantic filters and organize content amid increased complexity, but offered limited resources to seamlessly navigate the record as well as limited search capabilities. It also provided limited support to achieving clinical overview, with a participant describing it as being unable to envision the story of that patient's illness trajectory. In this cycle, we sought to fill these gaps by shifting our focus back to the clinical overview task and creating *ChartWalk*: a tool that allows healthcare workers to more seamlessly navigate large and very complex patient records when conducting a text-centered chart review.

With the awareness that this is a challenging problem to solve and that there can be significant diversity in chart review needs across different healthcare practitioners, we found it best to seek early feedback from a wide range of professionals via an iterative design process. This allowed us to validate decisions early on and reset course as needed.

We began our iterative design journey by establishing a set of initial design requirements based on the literature and our prior works. Since our participant pool in past work consisted largely of general practitioners, we sought to diversity perspectives early on and invited a number of mental health professionals (which is known to be a very text-heavy practice) for a round of informal feedback. That helped further consolidate design requirements before we began our first prototype. After this, we conducted two rounds of prototyping and evaluation which are described in Sections 5 and 6.

Throughout our design process we continuously recruited more participants, but also built a small network of collaborating professionals we had a chance to consult again later on, establishing some continuity to the design and evaluation process. We introduce all our participants first before delving deeper into our initial requirements.

4.1 Participants

We recruited a total of 19 healthcare practitioners from two institutions — a local pediatrics hospital (PH) and a local mental health hospital (MHH) — to take part in one or more of our three studies (Table 1). Three participants (P4, P8 and P9) took part in multiple sessions, which helped establish some continuity across iterations. Our recruitment criteria was *chart review complexity*: prospective participants' position either required them to do very in-depth and comprehensive chart reviews (i.e., most psychiatry and mental health related positions), or had a combined acute patient care situation dealing with large volumes and severe time constraints (i.e., critical care ICU). All recruited participants conducted chart reviews as part of their regular practice, albeit with slight variations in depth, priority and time pressure across specialties as demanded per their wide range of attributions.

4.2 Initial Design Requirements

Past work collectively mapped out lessons learned to support chart review workflows, based on assessments of physician practice [7, 40, 53, 60, 61]. Based on these, we compiled a list of preliminary design requirements **(DR_)** to kickstart the design process:

ID	Position	Org.	Interviews	1st Cycle	2nd Cycle
P1	Nurse Practitioner	PH	x		
P2	Psychiatrist	PH	x		
P3	Medical Resident	MHH	x		
P4	Resident/Psychiatrist	MHH	x	x	x
P5	Psychiatrist	MHH	x		
P6	Psychiatrist	MHH	x		
P7	Hospitalist	MHH	x		
P8	Behavioural Therapist	MHH		x	x
P9	Critical care Physician	PH		x	x
P10	Occupational Therapist	MHH		x	
P11	Psychiatrist	PH		x	
P12	Nurse	MHH		x	
P13	Nurse	MHH		x	
P14	Occupational Therapist	MHH		x	
P15	Social Worker	MHH		x	
P16	Nurse	MHH		x	
P17	Nurse	PH		x	
P18	Critical care Fellow	PH			x
P19	Critical care Physician	PH			x

Table 1. List of all study participants and what stages they contributed to.

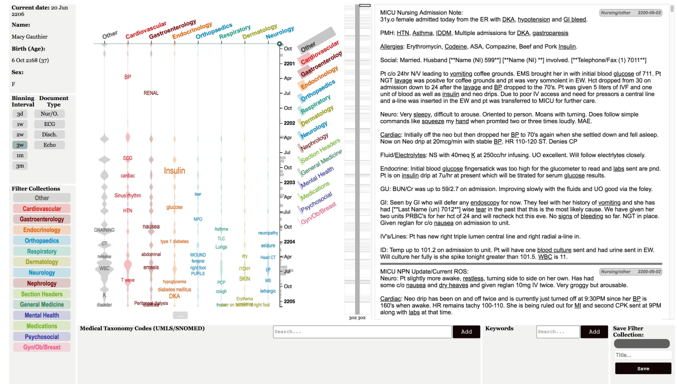


Fig. 1. Interface of Doccurate [61], presented in interviews for feedback. It features a panel with colored content filters mapped to medical sub-specialties (left), a timeline showing temporal frequency of terms under each colored filter (middle), and a view showcasing full notes (right).

→ **(DR1) Overview.** Provide high-level perspectives on the most prevalent issues about a patient, and offer starting points for exploration.

→ **(DR2) Time.** Support temporal awareness of medical events, when they happened, and how issues evolved.

→ **(DR3) Facets.** Organize information across meaningful content slices (i.e., subsets of related information, such as all mentions of "medications"), and provide means for users to stratify and locate content of interest.

→ **(DR4) Traceability and Context.** Support transitioning from the overview level to source notes via extensive linkage and preserve context of text excerpts when possible.

4.3 Initial Interviews

To contrast practices and consolidate our preliminary requirements (Section 4.2), we invited 7 healthcare workers (P1-P7) working in mental-health related positions (Table 1) across two different institutions for an informal discussion. We asked about their usual workflow and time frames when preparing to see a new patient, how their notes differed from those in non-mental health fields, what they expect to see in a summary, and any challenges they face navigating patient charts in their current EHR system. We also briefly introduced them to Doccurate [61] (Fig. 1), and asked their opinions on its graphics+text approach. Sessions were approximately 45 min long and participants received a $25 CAD gift card for their participation.

4.4 Findings

Participant practices were largely compatible with our preliminary understanding of the chart review process. We report these findings alongside the corresponding design requirements they complement.

Temporal views are appreciated (DR1, DR2) – When presented to Doccurate, most participants (5/7) appreciated the Timeline as an additional view to their study, commenting it was *"super helpful"* (P4), that it *"makes a lot of sense"* (P5), and that having a *"pictorial timeline"* would be more helpful than a text narrative (P4). This confirms past findings on the importance of temporal reasoning for chart review, and its benefits for overview purposes [60].

What goes in a summary (DR1, DR3) – Participants reported many challenges to find standard information from notes, especially on *"a time crunch"* (3/7). Medications (P3, P5), main health concerns (P4), legal issues (e.g. *"risk of suicide attempt"*) (P3, P6), and teasing out psychiatry-related details from non-mental health notes (P2) were a few examples of hard-to-find tidbits. This also matched information they reported expecting in a summary, which included medications (6/7), active problems (6/7), and social/family history (4/7).

Augment Search & Filter (DR3, DR4) – Participants reiterated a need for a keyword search function (P3, P6), and mentioned having a hard time to *"find the right note with the right information"*(P6) in their current practice. They mentioned it was also important to know *"which section that that word appears"* (P3), implying that context is important to retain.

Heavy reading loads, more time needed (DR4) – Most participants reported spending 15 to 20min perusing the record (min: ∼5 min (P1), max: ∼30 min (P2, P7)), which is above average compared to the general practice participants we interviewed in prior work (2-10min) [60]. While we could not authoritatively assess reasons for that (e.g., to what extent that is due to specialized information needs or due to their notes being reportedly *"lengthier"* (P1, P4) and nonstandard/hard to peruse (P3, P6)), we could expect an increased need for deeper dives into the source notes for this population. This calls for an emphasis on reading support and facilitated note retrieval and navigation.

5 1st Design Cycle: Consolidate Requirements

Based on the insights compiled in Section 4, we built an initial prototype as minimum viable solution to fulfill outlined requirements and conducted a round of interim evalutions to validate our design concept and refine design directions.

5.1 NLP and Data Processing

While patient records in healthcare practice encompass a mix of structured and unstructured data (such as ICU continuous body measurements), our focus was on information contained in text documents and how much information can be garnered from these sources in raw unstructured form. In line with this goal, we collected all text note collections from four large patient records in MIMIC-III [34], a large database of de-identified health records from critical care patients. Note collections contain 302, 551, 795 and 976 notes each, and were purposefully chosen for their size and complexity.

We processed the notes using the Google Healthcare Natural Language API [21], a named entity recognition (NER) service for medical text. The NER engine assigns multiple tags to detected entity mentions, including *type* (`PROBLEM`, `MEDICINE`, `LABORATORY_DATA`, and so on), *temporality* (whether a mention is cited as part of current issues, past issues, upcoming concerns, and so on), *certainty* (assesses terms are qualified or negated, e.g. *"no edema"*), a *subject* (whether the term is related to the patient or a family member), and *relations* (a qualified relation between entity mentions, e.g., between a `LABORATORY_DATA` mention and its corresponding `LAB_RESULT` mention and `LAB_UNIT` `mention`). Entity mentions may also be assigned a set of UMLS codes: the Unified Medical Language System [6] is a vast health and biomedical dictionary connecting many knowledge bases and terminologies.

Following, we processed entity mentions to organize them into meaningful groups. We first stratified mentions into type-based categories (*Problems*, *Medications*, *Labs*, *Family History*, *Social History*,

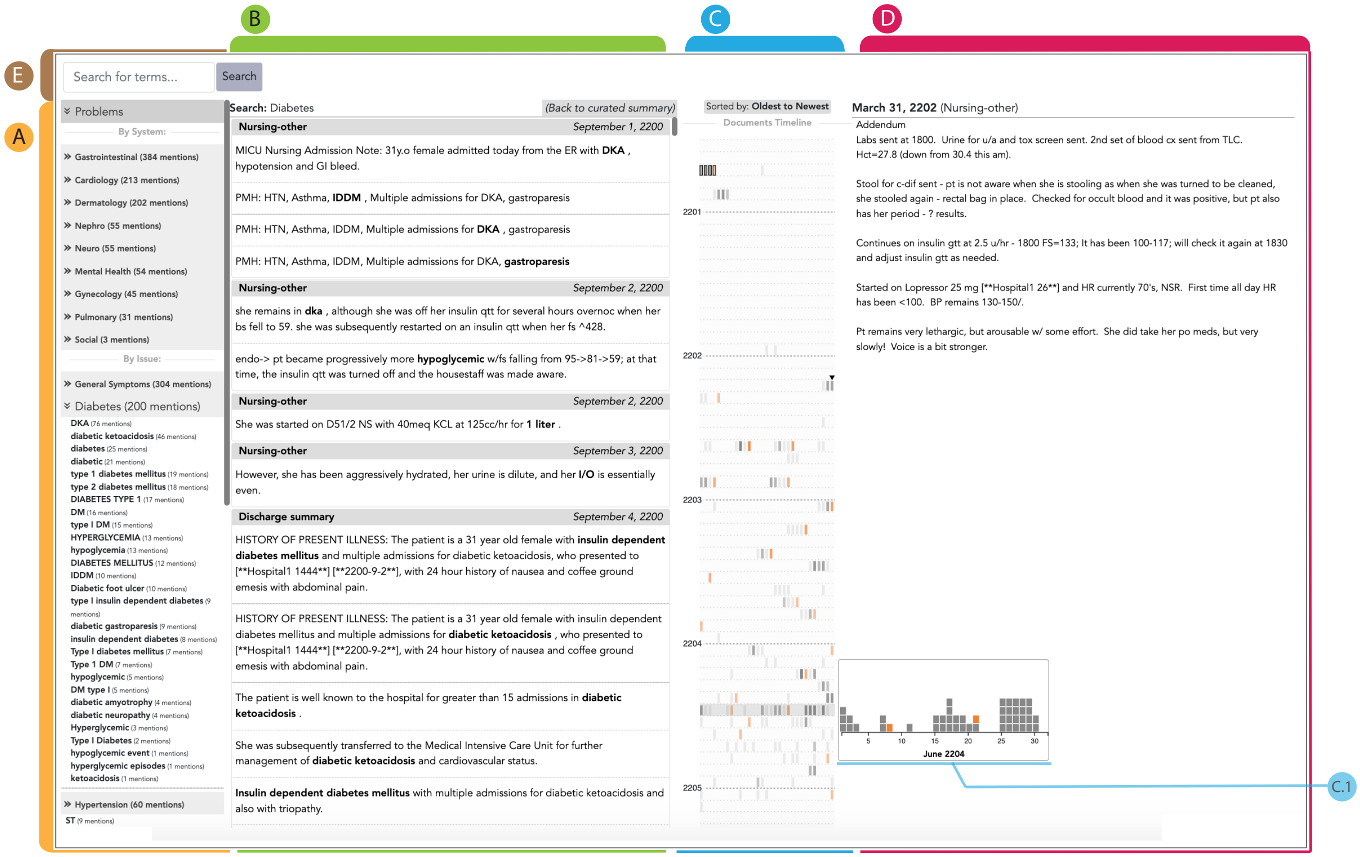


Fig. 2. Interim interface design of *ChartWalk*. (A) The *Mentions View* features an organized list of occurrences of clinically meaningful topics, organized by categories (e.g. Problem, Labs & Measurements) and sub-categories (e.g. Cardiology, Gynecology), sorted by frequency. (B) The *Snippets View* features segments of relevance from clinical notes, i.e., containing a mention from the mentions list, or a user-defined keyword entered in the search box (E), or previously curated segments. (C) The *Calendar View* provides a visual summary of the collection of notes over time, supporting temporal awareness and allowing quick navigation to notes. (D) The clinical notes of interest, pulled by clicking on a snippet in the Snippets view or a note marker in the calendar view, are displayed in the *Notes View*.

Allergies). Within each group, we further grouped mentions into *mention sets*, i.e., sets of the same concept across notes that accounted for alternative spellings. Mention sets were formed either via UMLS code overlap — mentions were merged if 2/3 of their UMLS codes matched — or via exact text match (case insensitive). While this strategy failed to merge some terms which were not UMLS-tagged by the NER engine, it still managed to significantly reduce the pool of mentions into meaningful sets. To further simplify the bigger pools of mention sets (i.e., *Problems* and *Medications*) type-based category lists, we further grouped frequent related mention sets into semantic sub-categories. Based on our findings from Doccurate [61], we curated both systems-level categorizations (e.g., "Cardiology", "Nephrology") and problems-based categorization (e.g., "Diabetes", "Hiperlipidemia") to encompass relevant frequent mentions.

5.2 Interface Design, in Context

The tool was created as a Javascript React and D3 application with a Python Flask server. The interface design comprises 5 main panels (Fig. 2): (A) a *Mentions view*, listing categorized mention sets **(DR1)**, (B) a *Snippets view* featuring two content modes (*"Search"* and *"Curated Summary"*) **(DR3, DR4)**, (C) a *Calendar view* as a timeline of all documents in the patient record **(DR2)**, (D) a *Note view* showing a full selected note, and (E) a *Search bar* **(DR4)**. We showcase how all views work together via a hypothetical chart review scenario.

Starting from the *Mentions view* (Fig. 2(A)), a physician conducting a chart review for a new patient would have access to a multi-level list of mention sets, organized under the type-based categories (and sub-categories if applicable) described in Section 5.1, and sorted by frequency so that more prevalent categories and mention sets appear on top. Infrequent and uncategorized mention sets (i.e., containing only 1-5 mentions) are further collapsed into sub-categories at the bottom to reduce clutter. The higher level type-based categories provides an overview of prevalent issues **(DR1)**, so the physician can skim over this list to get a general picture of this patient's health state and medical concerns they should be diving into next in their study. Under Problems, they spot that mentions of *"diabetes"* and a few other related issues (e.g., *"diabetic ketoacidosis"*) figures among the most frequent mention sets, indicating this is a major health concern for this patient.

They then have a look at the curated summary **(DR3)** in the *Snippets view* (Fig. 2(B)), containing snippets from various notes in the record **(DR3, DR4)**, pre-selected by a fellow nurse who had a chance to briefly visit this chart before. Snippets are grouped by note and presented chronologically from oldest to newest. Featured passages add important context to the diabetes issue, and the first few snippets indicate this patient is indeed struggling to control her diabetes while aggravated by other cardio and gastrointestinal problems, and that her family situation is difficult and contentious **(DR4)**.

Following, the physician also gets a quick look at the *Calendar view* **(DR1, DR2)** (Fig. 2(C)). This panel features a modified calendar format to encompass the whole documented history of this chart, in which every row represents a month and every colored marker in a row represents a day in that month that contains at least one note written for that patient. Orange day markers point to the presence of a discharge note on that day, and increased opacity maps to a higher number of notes available on that day. With a quick glance, the physician is able to identify periods of time with increased healthcare system utilization, which typically means a worsening health trajectory during that time **(DR1)**. Corresponding day markers for snippets currently visible in the *Snippets view* are further outlined in black which helps quickly situate the snippets in time, and get updated as the physician scrolls down the list **(DR2)**.

The physician identifies that there is an uptick in documentation in recent months, and opts to toggle sorting from "Oldest to Newest" to "Newest to Oldest", so that the more recent months in the *Calendar view* and most recent snippets in the *Snippets View* appear first. They then proceed to locate the latest discharge summary for a closer read. Upon finding the latest month with a orange day marker, they hover over the month row to bring up a side view displaying all notes written for that month (Fig. 2(C.1)). The physician locates the orange note marker to bring the discharge summary to the *Note view* **(DR4)** (Fig. 2(D)); a triangular indicator in the *Calendar view* now indicates the date of the currently selected full note.

After skimming through the latest discharge summary, the initial impressions about this patient are further consolidated, and the list of issues to learn more about has grown. The physician turns their attention back to the *Mentions view* for a dedicated look at each of these issues. After locating "Diabetes" on the list, the physician clicks on it to retrieve respective text mentions into the *Snippets view* **(DR3)**, which is now updated to show 'Search' results, with originating sentences retrieved along bolded mentions for full context **(DR4)**. The *Calendar view* also updates to visually mute any day markers not included in the search through fade out (Fig 2(C)), helping the physician assess how the issue has persisted over the documented history **(DR2)**. Going over the list of snippets, the physician spots a mention of "diabetic foot ulcer", which was missed in previously perused documents. The physician clicks on the snippet, redirecting them to the full note in the *Note view* for a closer follow-up **(DR4)**.

With 2 minutes left before seeing the patient, the physician opts to spend the rest of the time delving deeper into the patient's family situation. Using the search bar, they do a direct search for mentions of *"children"*; as they type *"chil"*, a dropdown menu displays possible mention-based search items included in the *Mentions view*, but the only mention available is *"chills"* (under Problems). The physicians goes ahead with a direct string search instead, and from the snippets they find conflicting mentions about the patient having either four or five children **(DR3)**. The physician makes a mental note to ask for clarifications, and gets ready to meet the patient in person.

5.3 Interim Evaluation

We recruited 11 healthcare workers (P4 and P8-P17, see Table 1) for quick feedback sessions to share opinions on our interim design. Our goal was to verify whether our design and envisioned features made sense to our target population, and to identify areas for improvement. Sessions were approximately 1h long, and took place via video conferencing with participants remotely accessing the prototype via their web browser and sharing their screen. Sessions consisted of a walkthrough of the tool, a 10min free-form chart review, and a follow-up discussion on their experience. Questions covered what they learned about the patient, pros and cons of the tool and its various features, how the tool compared to systems in current practice, and to what extent the tool let participants reach "sufficient overview". We also probed their thoughts on creating their own curated summaries if they had the chance, and if they would find that useful somehow. Sessions were transcribed and findings were derived from thematic analysis. Participants received a $30 CAD gift card for their participation.

Findings from our analysis allowed us deeper insight to refine our initial list of design requirements:

Refine Aggregation and Overview (DR1) – On one hand, the *Mentions view* was appreciated by many (6/11): several participants felt it was thorough (P8), helpful (P13, P16) and that it helped inform key aspects of the patient with less effort (P9, P16, P17), which suggests this is a useful feature for the task. On the other, some participants felt that the list was too long and overwhelming (3/11), which calls for further refinement of the semantic categories aimed at reducing complexity and overall list size.

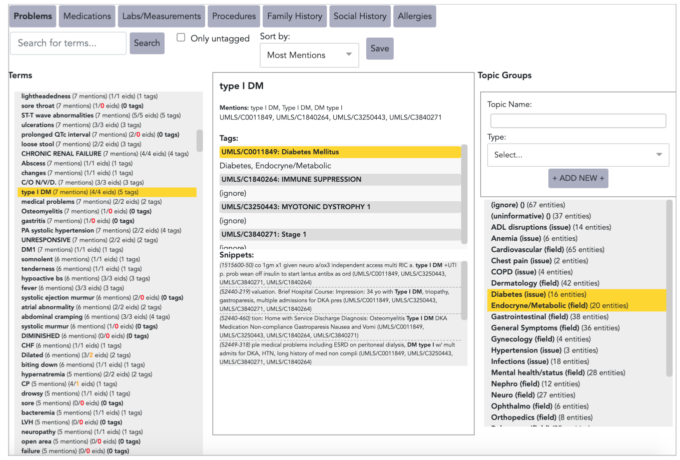


Fig. 3. The auxiliary tagger tool we created to help assign mentions (left, "Terms") to categories (right, "Topic Groups"). This is an administrative view not accessible to users.

Extract and display Time Series (DR2) – The *Calendar view* was also a well liked feature. While some participants found it overwhelming at first (3/11), several others found that the metaphor made sense (3/11), and that it was able to convey wellness patterns via healthcare utilization (i.e., period for which notes were written) (3/11). Participants appreciated the ability to flip the timeline to reverse chronological order, as this is what they use in their practice already and because it emphasizes content from the most recent notes (6/11): *"this is a better snapshot of right now, or the present time."* (P10). On the *Mentions view*, some concerns and suggestions were time-related. One participant missed the temporal context and how mentions are distributed over time, e.g.: *"did this happen once, or is it a recurrent issue?"* (P11). Other suggestions included using sparklines to relay frequency of mentions (P9) and access to related trending lab data and vitals (P4).

Readability and Navigability (DR3, DR4) – We intentionally refrained from using much color in our design, as at the time we did not know what data features would benefit the most from it. In the study however, participants who brought up color either appreciated being able to locate the discharge notes via color (2/11) or consistently mentioned they expected it to be used to depict note types (4/11); this makes sense, as each specialty regards some notes more useful than others and that assessment goes beyond the discharge summary. Other comments were about text, and that *ChartWalk* was still *"note heavy"*(P17), font *"could be bigger"* (P15), and that the snippets presentation could benefit from clearer separations, and less *"plain black text on white background"* (P14). These comments point not just to simple opportunities to improve readability, but also to a need to reconsider how much text gets displayed at a time and reassess surrounding visual cues.

Our interim evaluation also allowed us to identify a new design requirement to append to our list:

→ **DR5. Curation.** We came up with the idea to add the "Curated summary" based on our past work with curated filters in Doccurate [61]. We saw this as a synergistic way to leverage (curated) perspectives crafted by colleagues in the exercise of their own practice that could be useful to others, and we wanted to get their opinions if the concept made sense. In the end, participants not only appreciated the concept behind the "Curated Summary" (9/11) but also reported a myriad of use cases, including sharing content within their team (4/11), as a reminder to help them write their own notes (3/11) and when revisiting the record (2/11), and transferring care to another practitioner (P14). For those on the fence, one stated still finding it useful only if they are revisiting the record later (P11), or if it enables a very quick overview of issues (P17). Given this largely positive feedback, we felt this was worth pursuing further and included it as an additional design requirement.

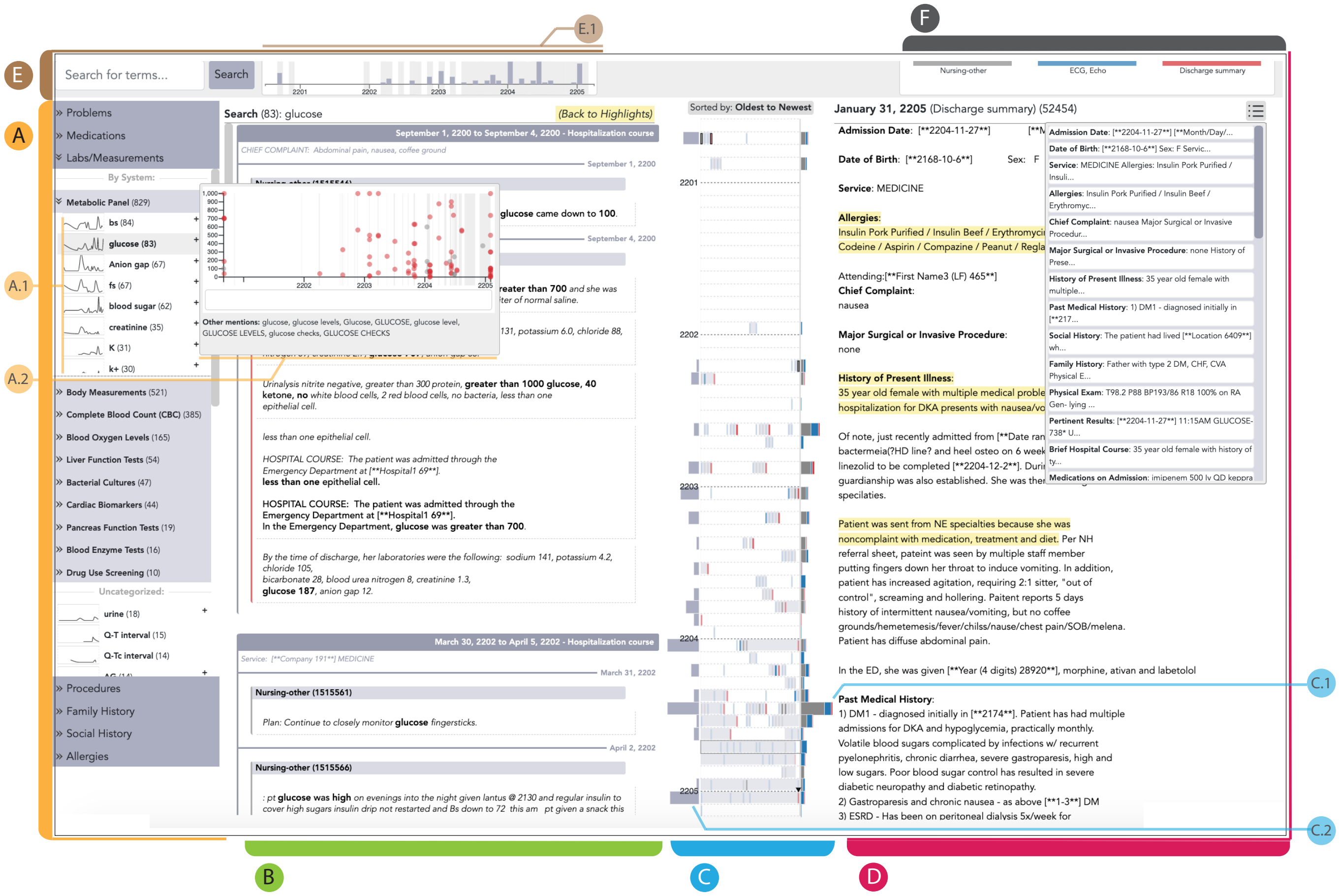


Fig. 4. Revised design of *ChartWalk*, comprising the same main panels as the interim design (see Fig. 2). (A) The *Mentions View* is further augmented with temporal support via sparklines (A.1) indicating temporal trends of mentions, and scatterplot pop-ups displaying automatically extracted numerical values from labs and measurements over time (A.2). (B) The *Snippets View* has been re-organized to better convey the hierarchical semantic context from which the snippets originate from, i.e. by note at the lowest level, then by date, then by episodes of care (e.g. hospitalization stay) at the highest level. (C) The *Calendar View* features two vertical histograms conveying the number of notes per row (i.e. month) (C.1.) and number of notes relevant to the currently active search (C.2.). (D) The Notes View is augmented with header formatting for readability, a document index for navigation, and also supports user-selected highlights, i.e. selection of segments of interest in a click to enable curation of information (shown in yellow). (E) The *Search Bar* encompasses a visual widget that conveys frequency of occurrences of the current term or topic of interest, similar to (C.2). Finally, (F) is an interactive *Legend Bar* that allows the clinician to filter out notes by type.

6 2ND DESIGN CYCLE: ASSESSING IMPACT ON WORKFLOWS

Based on the findings from the previous cycle, we redesigned *ChartWalk* [1] and expanded the set of features (Fig. 4). Our goal at this stage was to understand the impact of *ChartWalk* on the chart review process, which we assessed through five qualitative, in-depth task-based evaluation sessions. We discuss our findings on workflow and participant feedback, including opportunities for follow-up design cycles.

6.1 (Re)design and Development

The list of improvements below was largely informed by the findings in the 1st design cycle, but also include a few small usability additions made after the first two evaluation sessions in this 2nd design cycle (i.e., note headers, filtering by note type). The additions are as follows:

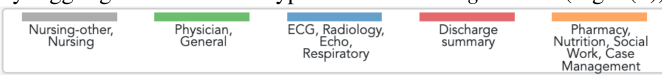
Expand categories in the *Mentions view*. Given the interest and reliance on the *Mentions view*, we chose to further expand and curate semantic categories. To help with this task, we devised an auxiliary tagging tool (Fig. 3) to help create new categories and associate them to mention sets (via UMLS codes or labels). At this stage, mention assignment is done on a one-by-one basis, although stemming, plus an ontology-based umbrella approach [61] would potentially expedite the curation process and would be fully compatible with this strategy. Publicly available medical knowledge sources were used to curate semantic categories for medications [18] and procedures [47], and we further expanded the systems-level categorizations for problems to include more specialties. We also added a new "Procedures" type-based category, for a total of seven high-level categories in the *Mentions view*.

Introduce episodes of care. We introduced the notion of *episodes of care* to simplify and further aggregate content pieces. In this context, episodes of care refer to periods of continued medical attention (e.g., as part of a single hospitalization course) and notes within an episode of care tend to have shared context and purpose. To detect episode time spans, we first extracted admission and discharge dates for hospitalization courses available in discharge summaries, and then generated spans from consecutive notes that did not fit any hospitalization course. The *Snippets view* was redesigned to further group documents into episodes of care, and now also displays the chief complaint (i.e., reason for hospitalization) when available (Fig. 4(B)).

Stratify content by note type. We also made it possible to discern more note types beyond discharge summaries at a glance, via color. Among our four patient record collections, we found a total of 13 different note types, some redundant (e.g., "Nursing-other" and "Nursing"). To keep the number of colors to a manageable level, similar note types were clustered into 5 groups (see figure below): nursing (grey), physi-

[1] This latest version of *ChartWalk* is available at `http://chartwalk.cs.toronto.edu/`,

cian notes (green), medical examination reports (blue), discharge (red) and others (orange). A color legend was added to the top bar to convey the groupings (Fig. 4(F)) appearing in a note. On the *Calendar view*, we added a horizontal stacked bar chart to each month to convey the number of notes of each group type (Fig. 4(C.1)), which precludes the need to hover over the months to see what note types are present within. Day markers also display the color of the least frequent note type for the day, making it easier to locate said notes. After the sessions with the first two participants in this cycle (P4 and P18), we felt it was important to further increase discoverability of notes and added the ability to filter items in the *Calendar view* and the *Snippets view* by type; this is done by toggling individual note type labels on the *Legend Bar* (Fig. 4(F)).



Improve temporal awareness. Following participant feedback from the previous cycle, we included a sparkline widget to mention sets in the *Mentions view* (Fig. 4(A.1)) to indicate how term frequency changed over time. The sparklines help convey when the mentions first emerged, when they were most prevalent, and whether they are likely to still be an active concern. For mention sets with associated numerical values (as per the identified *relations* between mentions described in Section 5.1), we display them as a scatterplot that pops up when hovering the mention set (Fig. 4(A.2)). Scatterplot items are colored by respective source note type, and allow for individual values to be inspected on hover; clicking on a scatterplot item will retrieve the mention in the *Notes view*. Our third temporal addition was a bar chart view to showcase distribution of search snippets over time. It is displayed beside the search bar for overview (Fig. 4(E.1)), as well as a bar chart matching the months in the *Calendar view* (Fig. 4(C.2)). This mirrored bar chart setup around the *Calendar view* (Fig. 4(C.1-2)) allows for search prevalence to be more easily compared against note distribution, to assess how search term peaks match note peaks.

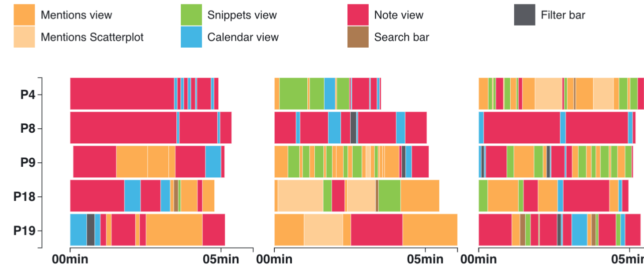
Enable user curation via highlighting. Following positive reactions to the idea of curating sets of relevant snippets, we included the ability to add and remove note content as personal *highlights* **(DR5)**. We streamlined the process as much as possible to allow highlights to be added and removed with a single click, with the trade-off that a full sentence gets highlighted instead of a custom text range. Highlighted segments appear on the *Note view* with a yellow background, and are displayed in the *Snippets view* in the space originally dedicated to the "Curated Summary".

Add skimming aids. After the sessions with the first two participants in this cycle (P4 and P18), we opted to add a few accents to help users navigate long notes in the *Note view*. We used a simple strategy to detect section headers via regular expressions, now rendered in boldface. We also added a navigable document index, listing shortcuts to all sections in the note (Fig. 4(D)).

6.2 Task-based Evaluation

For this 2nd design cycle we recruited 5 healthcare professionals — including 3 past (P4, P8, P9) and 2 new (P18 and P19) participants — to perform a chart review using the new version of *ChartWalk*, followed by a discussion on their experience and how it compared to their current practice. Similar to the 1st design cycle, participants were instructed to *"use the tool to the fullest, in a way that makes sense to you and your workflow"*, but were allotted more time to do the chart review (15min) and were asked to do mini-reports every 5min on what they learned and what remained to be learned. This allowed us to capture how information was prioritized and progressively refined throughout the process. Participants were also asked to highlight passages in notes that they found relevant, and were encouraged to refer back to them for their mini-reports. Finally, they were encouraged to think aloud if they felt comfortable doing so. Participants received a $40 CAD gift card for their participation.

We collected screen and audio recordings of the chart review exercise and the follow-up discussion. The chart review segments were timestamped and coded to indicate what activity was estimated to be performed at the time, and the discussions were transcribed and analyzed via thematic analysis. Following, we report our findings on common usage patterns, relevant participant impressions on the tool, and opportunities for improvement.



ID	Note	Mentions	Snippets	Calendar	Scatterplot	Search	Filter
P4	**43.5%**	17%	21%	7%	11%	0.5%	*
P8	**88.5%**	0%	0%	10%	0%	0%	1.5%
P9	30%	**38%**	22%	7%	1%	0%	2%
P18	**35%**	30%	10%	7.5%	16%	1.5%	*
P19	**42.5%**	34%	2.5%	9%	7.5%	2%	2.5%

** feature not available at the time*

Fig. 5. **Top**: Workflow charts depicting what views participants spent their time on at each 5 min chart review segment. Differences in segment length are due to small variations in manual time keeping; P4's second session ended earlier due to a time keeping error. **Bottom**: Aggregated percentages per participant on time spent in each view.

6.3 Findings on Workflow

We discuss major findings on how participants conducted their chart review, and how the tool contributed or hindered the task.

6.3.1 Diverse Strategies

Prior work revealed that chart review workflows encompass various tasks (Section 2). Usage patterns depicted in Fig. 5 reveals such diversity reflected on an individual level, showcasing a variety of workflows. While P4 and P8 chose to spend most of their time doing deep dives and reading past notes, P9, P18 and P19 split their time between deep dives and broader exploration and overview. This may be partly attributed to specialty and past EHR experience, as both P4 and P8 work in mental health and work in the same institution, whereas P9, P18 and P19 work in critical care under limited time frames and are used to operating on broader overviews. There is also a novelty factor to consider, as the new participants touched on the friction of learning this new tool, citing that *"it is a little overwhelming at first"* (P19) and *"I had to focus my energy on where I could find information"* (P18). Given time, we posit that *ChartWalk*'s ability to support a variety of workflows can help users pace themselves better when climbing this learning curve, and be compatible with a variety of personal chart review styles.

6.3.2 Reading (still) comes first

Despite the variety in workflows, a common (and unsurprising) action across all participants was beginning their study with at least one deep dive into the *latest* discharge summary. We argue this is more than just a residual habit: documentation like discharge summaries are crafted to be self-contained and therefore encompass most relevant things to know about a patient in a cohesive, fairly synthesized form. On the other hand, it does not always provide sufficient clinical overview since it can be *"very variable in its quality"*(P9), and participants appreciated *ChartWalk*'s overview features to help confirm, consolidate, and expand on these initial findings. That said, the takeaway is that reading remains an important task to perform, and that additional support to reading and skimming is instrumental to improving the chart review process.

6.3.3 Impressions on Information Gain

While we did not conduct any systematic performance assessment in this study, we did ask participants to compare how much they learned with *ChartWalk* versus what they would expect to learn with tools in their practice for the same amount of time. Their responses were varied, and not easily attributed to one particular feature. For example, while both P8 and P9 stated they learned more with *ChartWalk* than with tools in their practice, their workflows were wildly different: P8 prioritized deep dives, whereas P9 heavily relied on exploration of mentions. They each attributed this information gain to different features as well: the ability to quickly locate and triage notes by type via the timeline (P8) and overview capabilities via the *Mentions view* and the sparklines (P9). The other participants said they either learned about the same or a little less, which they partly attributed to friction with using a new tool, e.g., *"I think I learned a little bit less in the beginning, and I think partially that's influenced by being new and not knowing exactly where to find [things]"* (P19), and *"you develop all these workflows and workarounds such that when I'm doing a chart review for a patient I actually have a pretty consistent order that I click things"* (P18). In summary, while further studies are needed to properly ascertain information gain, reports suggest that a longitudinal validation approach may be needed to mitigate confounding effects related to this learning curve: *"We all internally created ways that we make sense of large amounts of data and our clinical patient world. We struggle when we are brought out of that process, that we've used for decades"* (P19).

6.4 Impressions on Tool

ChartWalk provided many new features that commercial EHRs do not, so we expected positive feedback despite the learning curve. More importantly however, participants detailed how these features made a difference in their workflow, which informs unmet needs in their current practice – some which strongly correlate with our initial design goals. We organize these findings under themes and opportunities for improvement in the following.

6.4.1 Global Views

Participants appreciated the different forms of global overview that *ChartWalk* provided. For the *Calendar view*, participants reiterated how it informed overall healthcare utilization and overall status of health (P19), but also, how the color coding and per-month histograms informed the kinds of healthcare visits during denser documentation periods (P19) and quickly assess if those are concerning or relevant to the study based on the types of notes (P18). The *Mentions view* was again appreciated for aggregating relevant items together (e.g. *"all medications"* (P4)), but more importantly, a perspective on whether issues were improving/worsening or active/resolved based on sparkline trends (P4, P9). The more granular grouping of issues plus sparklines also allowed for easier correlations between mentions, such as interactions between medical health and mental health symptoms (P9) and reflecting on prescription choices based on mentions frequency of two medication alternatives (P8). Participants acknowledged the dearth of such resources in their current EHRs, and that they are often unable to make such assessments in a timely manner.

Opportunity: Special focus on recent and active events. While *ChartWalk* provided means to capture the whole documented history of a patient, participants mentioned other status features that would be useful to access at-a-glance, such as discerning between active vs. resolved problems (P19), current vs. past medications (P18), and scheduled vs. *"as needed"* medications (P19). P18 also reiterated the need to emphasize recent events: *"As much as I dislike the EHR, it displays the labs and most recent vital signs pretty clearly. There was a time-specific data, 'what were the labs that were taken this morning' that the EHR does quite well and to be honest, an application like that just makes it much harder for me"* (P18).

Opportunity: Sharing across institutions. When explaining how different institutions (who do not share the same EHR infrastructure) share information via large printouts of notes as static reports, a participant suggested it would be useful if some overview components like the ones in *ChartWalk* could be appended to said reports: *"having a consolidated view like this about these trajectory views would rapidly accelerate the process of understanding what happened outside"* (P9).

6.4.2 Fluidly transition across perspectives

Participants appreciated the navigation tools available. Regarding the timeline, *"it's very nice to start out with the fifty-thousand foot view, and once you get a big picture, taking a deeper dive into what seems to be the most important things"* (P9). Relevant instances include being able to quickly zero in from the *Calendar view* to the relevant notes in denser time periods (P8) as well as quickly locating relevant notes by type (P4); it is worth noting that filtering by note type was used by all participants who had access to this feature (P8, P9, P19). Participants who successfully incorporated the *Mentions view* into their workflow also often chose to dive deeper and access the snippets for added context (P9, P18, P19).

Opportunity: Zooming in and finding relations. Suggestions to extend exploration features in *ChartWalk* included the ability to narrow down to smaller periods of time and extract localized relevant events (and respective documents) (P18), thus revealing details that get potentially drowned under the global history. To facilitate the detection of correlations from mention sparklines, another suggestion was to *"pick a key set of concerns, and just look at how the time series looked in a single overlaying figure"* (P9), which could be achieved via a new dedicated view for selected mentions.

6.4.3 Curate to prioritize, recall, and expedite

Content highlighting was the most spontaneously and consistently praised feature in *ChartWalk*. While we explicitly asked participants to highlight content, all participants found value and purpose in the task. They cited many possible uses, including bookmarking relevant passages to revisit later (P18, P19), using highlights as a consolidated summary to help dictate their own note (P4), and as memorization aids during the chart review process itself: *"I find it sticks in my mind a bit more"* (P8).

Opportunity: Share perspectives. Some participants discussed the potential impact of sharing and leveraging highlights among colleagues to expedite the chart review process, by aiding discovery of *"high yield"* areas (P18) and even using it as *"primary communication"* between them and other clinicians (P9): *"If I was just jumping into this chart and I knew that expert clinicians had been looking at it before, and I had one minute to review the chart, I could just look through the highlights if another clinician had done it, and say, OK, I understand the patient's problems."* (P9). One implied aspect here is that it does matter *who* generates the curation. When inquired to what extent they expected their colleague's highlights or categories to align with their own, answers rested on *"it depends"*: on one hand, while *"certain aspects are pretty much universal"* (P19), the expectation was that close colleagues (P18) or those who shared a field of specialty (P19) would produce more useful curation: *"what a medical student or nurse would highlight is very different than what I would highlight"* (P19).

Opportunity: Curate categories. While acknowledging that the high-level type-based categories made sense and *"fit nicely within a (familiar) framework"*, P19 expressed wanting to correct or readjust items in the Mentions view, either to fix automation mistakes (e.g., merge *"bs"*, *"blood sugars"* and *"fingersticks"* under a single mention), or to reassign mentions to different sub-categories (e.g., "Aspirin" visible under "Analgesics", but likely prescribed for cardiovascular issues instead). This latter example — a medication with multiple therapeutic uses — highlights some of the challenges around medical NLP and underscores the certainty that mistakes will be made: this introduces overhead to the chart review workflow (e.g., *"it required more effort for me to go back and look and verify that was a mistake by the system, versus in the EHR, once somebody has purposefully entered something."* (P18)), which can be partly mitigated by making curation and error correction tools more widely available for users. We point to our past work, Doccurate [61], for a more in-depth discussion on category curation and user perception of errors in automation.

6.4.4 Skimming aids

Participants commented on text presentation and reading. Even when more closely reading a note, healthcare workers will still skim a note looking for the sections they need, which happened consistently in our study. On that regard, P8 appreciated the bolded headers, which helped quickly locate those sections and conferred some *"consistency"* to the unfamiliar note formats they perused. P4 and P9 also extensively used the snippets, which bypassed the need to visit individual notes and collectively informed on specific issues (even with limited context).

Opportunity: More skimming aids. Despite acknowledging that using *ChartWalk* was more efficient than their current practice, P9 found there is *"still more text than I find useful"* when perusing notes, and that *"they contain maybe 75% of information that you don't need"*. They proposed *"an ability to stratify quickly between positives and negatives"*, which could be achieved for example by reducing text opacity on negative mentions (e.g., *"no edema"*).

7 Discussion

We briefly reflect on the collective findings of our iterative design journey. We begin by sharing lessons learned to hopefully guide design next steps in supporting the chart review process in healthcare. We then discuss more general implications of our lessons learned as promising opportunities for future work that extend into other text analysis-oriented domains beyond chart review.

7.1 Lessons Learned

Reading to ground abstractions. Our findings show that reading is not only very prevalent in medical practice, but it is also likely to continue to be prevalent even as better automated content overviews — graphical summaries and abstractive text summarizations alike — are more widely available in healthcare information systems. Prior research we conducted [60] confirms the importance of reading to existing chart review workflows: not only does text enable a nuanced understanding of the patient's story, physicians have also been extensively trained on them, making reading an important practice to support. Apart from these factors, notes also play an important summarization and evidential role [30] and should continue to be accessible front and center if only as a trustworthy and fail-safe strategy to ensure nothing critical is being missed. We also posit this should have a positive effect in fostering adequate levels of trust and effective reliance on automated outcomes [27].

Investigate skimming strategies. Our observations and past work [46] have shown that a lot of interaction with text happens in the form of *skimming*, and so leveraging and supporting this task may lead to concrete efficiency gains in the chart review. We envision explorations in text highlighting [58], text as data encoding [8] and word-sized graphics [5] are good places to start towards that goal.

Further seize recent and active findings. Our findings in the 2nd design cycle included comments on how participants hoped to be able to see active and recent issues more clearly. While the sparklines in *ChartWalk* offer hints as to whether a mention is potentially a current concern, an immediate ramification would be to flag or sort such items in the *Mentions view*. However, this principle could be extended to other content segments as well. For example, timelines could be designed to be adaptive and allocate more real estate to recent periods of time (e.g., *à la* Powsner and Tufte's Graphical Summary [52]), and note views could render extra emphasis on mentions of active issues.

Expand curation (with caution). The highlighting feature, despite its relative simplicity, prompted enthusiastic reactions and was touted as a helpful function. We can definitely see interesting extensions being possible, such as selectively merging highlights from a large set of users to try and find more general-purpose views, or how collective highlights can be used to rank and prioritize content pieces in search and visual summaries. On the other hand, we do see more urgent open questions to consider first. First, while highlighting was deemed useful for both personal and collective purposes, the interplay between these two levels and the ideal middle-ground for collective highlights are not yet clear. Second, it is also unclear how or whether highlights should be managed as they "age", and the extra work it would entail if management is left solely in users hands. We believe these questions should be investigated first, possibly in tandem with automated approaches for highlighting (i.e., extractive summarization).

7.2 Promising research directions beyond chart review

Synergistic curation towards effective automation. The positive reception and perceived usefulness of highlights and other low-effort curation features is great news for medical text NLP, as motivated users leveraging features meaningful to their practice are more likely to generate higher numbers of quality labels — as opposed to framing it as an extra task to perform on top of their duties. Well established curation practices might also pave ground to more seamless integration of automation into current systems, as users are eased into automated outcomes via a familiar metaphor of colleagues providing highlights to a group.

Improved overview tools to affect medical documentation. Participants in our study commented on how *ChartWalk*'s features would help them craft their own notes which suggests that chart review features could be useful to integrate into charting tools, a concept validated by past and recent work [43, 65]. Beyond supporting documentation tasks, we wonder if content retrieval could have an effect on how notes are represented and formatted. A factor that contributes to the bloating of clinical notes is the expectation for notes to be self contained, which is influenced by how difficult it is to find information in the EHR; it would be worth investigating whether that redundancy could be partly eliminated in favour of more concise documentation with better accompanying overview features.

Mediate stories with patients. One participant suggested that the summary elements available in *ChartWalk* could be shared with other practitioners across different institutions. We posit that the same principle could apply to patients and caretakers, featuring similar or slightly adapted visual overviews of their own records. This could help mediate conversations around medical issues, as well as foster transparency in the clinician-patient relationship.

8 Conclusion

While EHRs offer unprecedented opportunities to support safe and high-quality healthcare, there are still many barriers to effectively leveraging massive amounts of semi-structured clinical notes. The automatic extraction and visual summarization of knowledge gathered from these notes is a promising avenue to support clinicians in reviewing and synthesizing information from a large collection of notes. Yet, the source text documentation will still remain an indispensable tool to healthcare practitioners.

In this work, we embrace the essential nature of these imperfect textual artifacts by promoting them back as first class citizens via a mixed **graphics+text** concept: from structured information extracted from the notes via medical NLP, we augment these in a visual analysis tool that combines interactive visual summaries to support information retrieval, targeted navigation from and to the notes, and easy marking of segments of interest. Via a two-stage iterative design and evaluation process, we developed *ChartWalk*, which embodies this concept. Our design study allowed us to identify the merits and potential of using graphical summaries at the service of navigating text, rather than as a substitute to the original notes. Our work is a step towards expanding our understanding of what kinds of features healthcare workers see value in, while shedding light on the possibilities of combining current technologies in NLP, visualization and interaction, which we hope will provide inspiration for future works in this area.

Acknowledgments

We thank our healthcare worker participants and anonymous reviewers for their feedback and commitment. This research has been approved by the Ethics Board at the University of Toronto, and has been partially supported by NSERC (Natural Sciences and Engineering Research Council of Canada).

References

[1] E. Alexander, J. Kohlmann, R. Valenza, M. Witmore, and M. Gleicher. Serendip: Topic model-driven visual exploration of text corpora. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 173–182. IEEE, 2014.

[2] B. G. Arndt, J. W. Beasley, M. D. Watkinson, J. L. Temte, W.-J. Tuan, C. A. Sinsky, and V. J. Gilchrist. Tethered to the ehr: primary care physician workload assessment using ehr event log data and time-motion observations. *The Annals of Family Medicine*, 15(5):419–426, 2017.

[3] J. W. Beasley, T. B. Wetterneck, J. Temte, J. A. Lapin, P. Smith, A. J. Rivera-Rodriguez, and B.-T. Karsh. Information chaos in primary care: implications for physician performance and patient safety. *The Journal of the American Board of Family Medicine*, 24(6):745–751, 2011.

[4] F. Beck, S. Koch, and D. Weiskopf. Visual analysis and dissemination of scientific literature collections with survis. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):180–189, 2015.

[5] F. Beck and D. Weiskopf. Word-sized graphics for scientific texts. *IEEE Transactions on Visualization and Computer Graphics*, 23(6):1576–1587, 2017.

[6] O. Bodenreider. The unified medical language system (umls): integrating biomedical terminology. *Nucleic acids research*, 32(suppl_1):D267–D270, 2004.

[7] C. Bossen and L. G. Jensen. 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*, pp. 257–268. ACM, 2014. doi: 10.1145/2531602.2531620

[8] R. Brath. *Visualizing with Text*. CRC Press, 2020.

[9] M. Brehmer, S. Ingram, J. Stray, and T. Munzner. Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *IEEE transactions on visualization and computer graphics*, 20(12):2271–2280, 2014.

[10] A. A. Bui, D. R. Aberle, and H. Kangarloo. Timeline: visualizing integrated patient records. *IEEE Transactions on Information Technology in Biomedicine*, 11(4):462–473, 2007. doi: 10.1109/titb.2006.884365

[11] M. Correll, M. Witmore, and M. Gleicher. Exploring collections of tagged text for literary scholarship. In *Computer Graphics Forum*, vol. 30, pp. 731–740. Wiley Online Library, 2011.

[12] F. Dabek, E. Jimenez, and J. J. Caban. A timeline-based framework for aggregating and summarizing electronic health records. In *2017 IEEE workshop on visual analytics in healthcare (VAHC)*, pp. 55–61. IEEE, 2017.

[13] N. Diakopoulos, M. Naaman, and F. Kivran-Swaine. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *2010 IEEE Symposium on Visual Analytics Science and Technology*, pp. 115–122. IEEE, 2010.

[14] M. Dörk, D. Gruen, C. Williamson, and S. Carpendale. A visual backchannel for large-scale events. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1129–1138, 2010.

[15] W. Dou, X. Wang, D. Skau, W. Ribarsky, and M. X. Zhou. LeadLine: Interactive visual analysis of text data through event identification and exploration. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pp. 93–102. IEEE, 2012.

[16] J. C. Feblowitz, A. Wright, H. Singh, L. Samal, and D. F. Sittig. Summarization of clinical information: a conceptual model. *Journal of biomedical informatics*, 44(4):688–699, 2011.

[17] J. C. Feblowitz, A. Wright, H. Singh, L. Samal, and D. F. Sittig. Summarization of clinical information: a conceptual model. *Journal of biomedical informatics*, 44(4):688–699, 2011.

[18] U. Food and D. A. (FDA). General drug categories, 2015.

[19] S. Fu, J. Zhao, H. F. Cheng, H. Zhu, and J. Marlow. T-cal: Understanding team conversational data with calendar-based visualization. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2018.

[20] B. S. Glicksberg, B. Oskotsky, P. M. Thangaraj, N. Giangreco, M. A. Badgeley, K. W. Johnson, D. Datta, V. A. Rudrapatna, N. Rappoport, M. M. Shervey, et al. Patientexplorer: an extensible application for dynamic visualization of patient clinical history from electronic health records in the omop common data model. *Bioinformatics*, 35(21):4515–4518, 2019.

[21] Google. Healthcare natural language api.

[22] C. Hallett. Multi-modal presentation of medical histories. In *Proceedings of the 13th international conference on Intelligent user interfaces*, pp. 80–89. ACM, 2008. doi: 10.1145/1378773.1378785

[23] C. Hallett, R. Power, and D. Scott. Summarisation and visualisation of e-health data repositories. In *UK E-Science All-Hands Meeting, 18-21 Sept 2006, Nottingham, UK*, 2006.

[24] A. Handler and B. O'Connor. Rookie: A unique approach for exploring news archives. *arXiv preprint arXiv:1708.01944*, 2017.

[25] J. S. Hirsch, J. S. Tanenbaum, S. L. Gorman, C. Liu, E. Schmitz, D. Hashorva, A. Ervits, D. Vawdrey, M. Sturm, and N. Elhadad. Harvest, a longitudinal patient record summarizer. *Journal of the American Medical Informatics Association*, pp. amiajnl–2014, 2014. doi: 10.1136/amiajnl-2014-002945

[26] P. Hobbs. The use of evidentiality in physicians' progress notes. *Discourse Studies*, 5(4):451–478, 2003. doi: 10.1177/14614456030054001

[27] R. R. Hoffman, M. Johnson, J. M. Bradshaw, and A. Underbrink. Trust in automation. *IEEE Intelligent Systems*, 28(1):84–88, 2013.

[28] W. Hsu, R. K. Taira, S. El-Saden, H. Kangarloo, and A. A. Bui. Context-based electronic health record: toward patient specific healthcare. *IEEE Transactions on information technology in biomedicine*, 16(2):228–234, 2012. doi: 10.1109/titb.2012.2186149

[29] M. Hu, H. Yang, M. X. Zhou, L. Gou, Y. Li, and E. Haber. Opinionblocks: a crowd-powered, self-improving interactive visual analytic system for understanding opinion text. In *Human-Computer Interaction–INTERACT 2013*, pp. 116–134. Springer, 2013.

[30] K. M. Hunter. *Doctors' stories: The narrative structure of medical knowledge*. Princeton University Press, 1991.

[31] E. Isaacs, K. Damico, S. Ahern, E. Bart, and M. Singhal. Footprints: A visual search tool that supports discovery and coverage tracking. *IEEE transactions on visualization and computer graphics*, 20(12):1793–1802, 2014.

[32] S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann. On close and distant reading in digital humanities: A survey and future challenges. In *Eurographics Conference on Visualization (EuroVis)-STARs. The Eurographics Association*, vol. 2, p. 6, 2015.

[33] S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann. Visual text analysis in digital humanities. In *Computer Graphics Forum*, vol. 36, pp. 226–250. Wiley Online Library, 2017.

[34] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

[35] S. Koch, M. John, M. Worner, A. Muller, and T. Ertl. VarifocalReader – in-depth visual analysis of large text documents. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):1723–1732, 2014.

[36] K. Kucher and A. Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In *2015 IEEE Pacific visualization symposium (pacificVis)*, pp. 117–121. IEEE, 2015.

[37] S. Liu, Y. Chen, H. Wei, J. Yang, K. Zhou, and S. M. Drucker. Exploring topical lead-lag across corpora. *Knowledge and Data Engineering, IEEE Transactions on*, 27(1):115–129, 2015.

[38] S. Liu, X. Wang, C. Collins, W. Dou, F. Ouyang, M. El-Assady, L. Jiang, and D. A. Keim. Bridging text visualization and mining: A task-driven survey. *IEEE transactions on visualization and computer graphics*, 25(7):2482–2504, 2018.

[39] S. Liu, M. X. Zhou, S. Pan, Y. Song, W. Qian, W. Cai, and X. Lian. Tiara: Interactive, topic-based visual text summarization and analysis. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(2):25, 2012.

[40] L. Mamykina, D. K. Vawdrey, P. D. Stetson, K. Zheng, and G. Hripcsak. Clinical documentation: composition or synthesis? *Journal of the American Medical Informatics Association*, 19(6):1025–1031, 2012.

[41] T. Mønsted, M. C. Reddy, and J. P. Bansler. 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*, pp. 81–100. Springer, 2011. doi: 10.1007/978-0-85729-913-0_5

[42] T. Munzner. *Visualization analysis and design*. CRC press, 2014.

[43] L. Murray, D. Gopinath, M. Agrawal, S. Horng, D. Sontag, and D. R. Karger. Medknowts: Unified documentation and information retrieval for electronic health records. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pp. 1169–1183, 2021.

[44] S. Narayan and C. Cheshire. Not too long to read: The tldr interface for exploring and navigating large-scale discussion spaces. In *2010 43rd Hawaii International Conference on System Sciences*, pp. 1–10. IEEE, 2010.

[45] E. Nygren and P. Henriksson. Reading the medical record. i. analysis of physician's ways of reading the medical record. *Computer methods and programs in biomedicine*, 39(1):1–12, 1992.

[46] E. Nygren and P. Henriksson. Reading the medical record. i. analysis of physician's ways of reading the medical record. *Computer methods and programs in biomedicine*, 39(1):1–12, 1992.

[47] N. I. of Health (NIH). Medical tests, 2022.

[48] H. Park and J. Choi. V-model: a new innovative model to chronologically

visualize narrative clinical texts. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 453–462. ACM, 2012.

[49] R. Pivovarov and N. Elhadad. Automated methods for the summarization of electronic health records. *Journal of the American Medical Informatics Association*, 22(5):938–947, 2015. doi: 10.1093/jamia/ocv032

[50] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman. Lifelines: visualizing personal histories. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 221–227. ACM, 1996. doi: 10.1145/238386.238493

[51] C. Plaisant, R. Mushlin, A. Snyder, J. Li, D. Heller, and B. Shneiderman. Lifelines: using visualization to enhance navigation and analysis of patient records. In *Proceedings of the AMIA Symposium*, p. 76. American Medical Informatics Association, 1998. doi: 10.1016/b978-155860915-0/50038-x

[52] S. M. Powsner and E. R. Tufte. Graphical summary of patient status. *The Lancet*, 344(8919):386–389, 1994.

[53] D. Reichert, D. Kaufman, B. Bloxham, H. Chase, N. Elhadad, et al. Cognitive analysis of the summarization of longitudinal patient records. In *AMIA annu symp proc*, vol. 2010, pp. 667–671, 2010.

[54] A. Rind, T. D. Wang, W. Aigner, S. Miksch, K. Wongsuphasawat, C. Plaisant, B. Shneiderman, et al. Interactive information visualization to explore and query electronic health records. *Foundations and Trends® in Human–Computer Interaction*, 5(3):207–298, 2013.

[55] S. T. Rosenbloom, J. C. Denny, H. Xu, N. Lorenzi, W. W. Stead, and K. B. Johnson. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *Journal of the American Medical Informatics Association*, 18(2):181–186, 2011.

[56] S. Sarkar and D. Seshadri. Conducting record review studies in clinical practice. *Journal of clinical and diagnostic research: JCDR*, 8(9):JG01, 2014.

[57] J. Stasko, C. Görg, and Z. Liu. Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization*, 7(2):118–132, 2008.

[58] H. Strobelt, D. Oelke, B. C. Kwon, T. Schreck, and H. Pfister. Guidelines for effective usage of text highlighting techniques. *IEEE transactions on visualization and computer graphics*, 22(1):489–498, 2015.

[59] N. Sultanum, A. Bezerianos, and F. Chevalier. Text visualization and close reading for journalism with storifier. In *2021 IEEE Visualization Conference (VIS)*, pp. 186–190. IEEE, 2021.

[60] N. Sultanum, M. Brudno, D. Wigdor, and F. Chevalier. More text please! understanding and supporting the use of visualization for clinical text overview. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 422. ACM, 2018.

[61] N. Sultanum, D. Singh, M. Brudno, and F. Chevalier. Doccurate: A curation-based approach for clinical text visualization. *IEEE transactions on visualization and computer graphics*, 25(1):142–151, 2019.

[62] X. Q. Tabassum Kakar, E. Rundensteiner, L. Harrison, S. Sahoo, S. De, and T. La. Context: Supporting the pursuit and management of evidence in text-based reporting systems. In *17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022)*, vol. 3, pp. 38–50. SCITEPRESS, 2022.

[63] M. Vassar and H. Matthew. The retrospective chart review: important methodological considerations. *Journal of educational evaluation for health professions*, 10, 2013.

[64] Q. Wang and R. Laramee. Ehr star: The state-of-the-art in interactive ehr visualization. In *Computer Graphics Forum*. Wiley Online Library, 2021.

[65] L. Wilcox, J. Lu, J. Lai, S. Feiner, and D. Jordan. Physician-driven management of patient progress notes in an intensive care unit. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1879–1888, 2010.

[66] K.-P. Yee and M. Hearst. Content-centered discussion mapping. *Online Deliberation*, 2005.

[67] J. Zhao, L. Gou, F. Wang, and M. Zhou. Pearl: An interactive visual analytic tool for understanding personal emotion style derived from social media. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on*, pp. 203–212. IEEE, 2014.