Visual Information Seeking in Multiple Electronic Health Records: Design Recommendations and A Process Model

Taowei David Wang
HCIL
2117 Hornbake Building
University of Maryland
College Park, MD 20742
tw7@cs.umd.edu

Catherine Plaisant
HCIL
2117 Hornbake Building
University of Maryland
College Park, MD 20742
plaisant@cs.umd.edu

Ben Shneiderman
Dept. of Computer Science
A.V. Williams Building
University of Maryland
College Park, MD 20742
ben@cs.umd.edu

ABSTRACT

In the advent of electronic health record (EHR) systems, physicians and clinical researchers enjoy the ease of storage, retrieval, persistence, and sharing of patient data. However, the way physicians interact with EHRs has not changed much. More specifically, task support for temporally analyzing large number of EHRs has been lacking. A number of information visualization techniques have been proposed to alleviate this problem. Unfortunately, due to their limited application to a single case study, the results are often difficult to generalize across medical scenarios. In this paper we present the usage data of and user comments on our information visualization tool Lifelines2 through eight different medical case studies. We generalize our experience into an information- seeking process model for multiple EHRs. Base on our analysis, we make recommendations to future information visualization designers for EHRs on common design requirements and future research directions.

Categories and Subject Descriptors

H.5.2 [**Information interfaces and Presentation**]: User Interfaces; H.1.2 [**Information System**]: User/Machine Systems—*Human factors*

General Terms

Information Visualization, HCI

Keywords

Information Visualization, Electronic Health Records, Design Requirements, HCI

1. INTRODUCTION

Whether it is to make diagnoses for a single patient, or to obtain quality assurance measures of health care by analyzing multiple patients, physicians and clinical researchers must incorporate large amount of multidimensional historic data. Electronic health record (EHR) systems facilitate the storage, retrieval, persistence, and sharing of patient health information. However, the availability of information does not necessarily translate to adequate support for complex tasks physicians and clinical researchers face everyday.

Overwhelmingly large amount of information and a lack of support for temporal queries and analyses are but a few problems physicians and clinical researchers face. A number of information visualization systems have been introduced to address these issues. These systems support higher-level decision-making and exploratory analysis tasks in the medical domain. Commendably, these systems aim to solve real problems physicians face, and add value to the EHRs for the end-users. However, these systems are often designed for one specific scenario, and subsequently evaluated on that scenario. As a result, it is difficult to make generalizations on how physicians seek information or on the user requirements for information-seeking in EHRs.

In contrast, our information visualization tool, Lifelines2, has been applied to eleven different case studies, eight of which are in the medical domain. Since Lifelines2's inception, we have worked closely with physicians and hospital administrators to gather user requirements for the tasks of search and exploratory analysis of multiple patient records over time. It has been used by physicians for the purpose of obtaining quality assurance measures, assessing impact on patient care due to hospital protocol changes, replicating published clinical studies using in-hospital data, and simply searching for patients with interesting medical event patterns.

Over the two and half year period in which these case studies took place, we observed how physicians used Lifelines2, logged the interactions and features physicians used, and collected physicians' comments. By analyzing the usage and user-feedback data, we were able to make generalizations about searching for information in EHRs. Section 2 first presents related work. We then present Lifelines2 in Section 3 describe one case study in detail. Following that, we present an analysis of Lifelines2 usage log data and a process model. Finally, we close with a list of design recommendations and concluding remarks.

2. RELATED WORK

As EHR systems become more prevalent, the need to develop appropriate techniques for users to interact with EHRs also become more pressing. A growing number of recent field

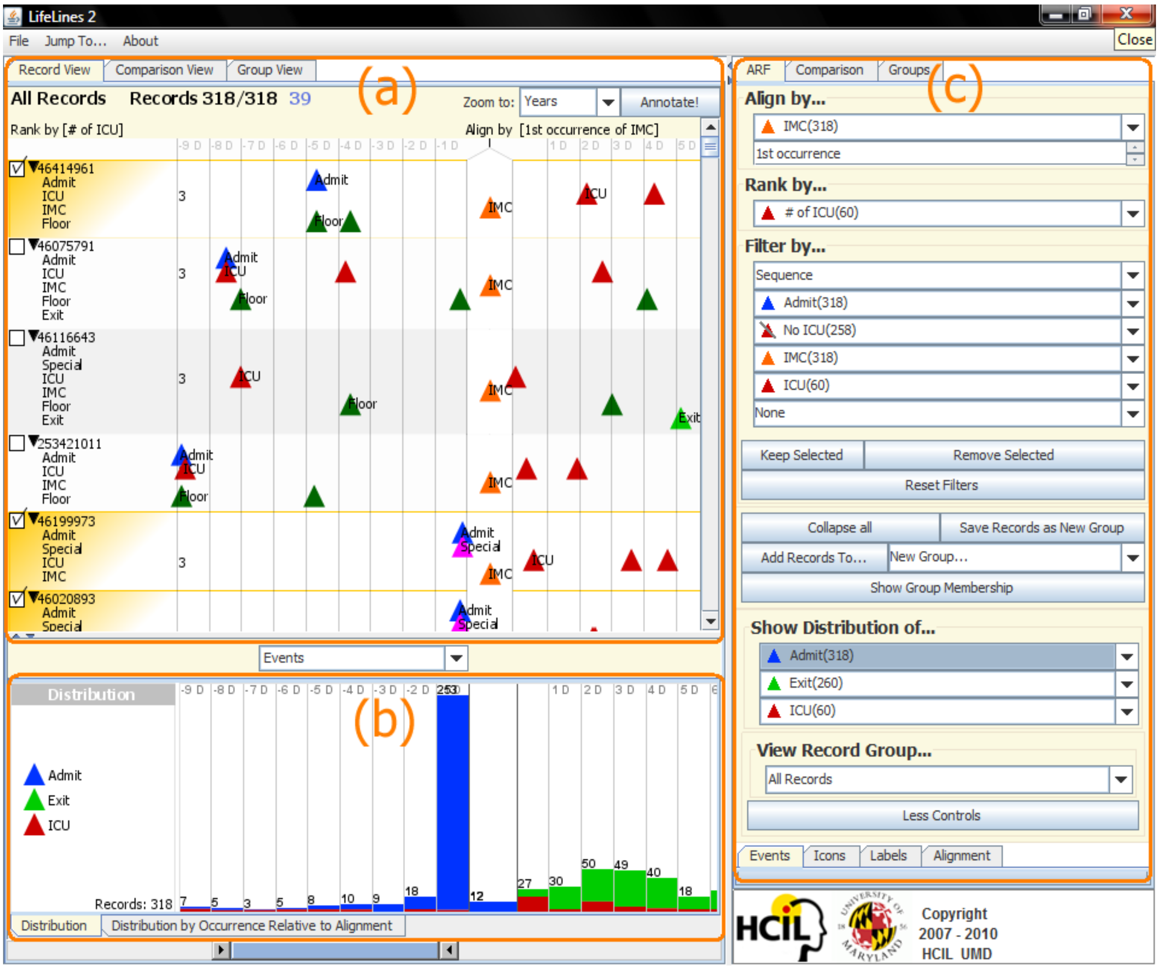

Figure 1: A screen shot of Lifelines2. (a) shows the main visualization of multiple EHRs. (b) is a temporal summary, showing the distribution of the three event types *Admit*, *Exit*, and *ICU* over time. (c) shows the control panel for Lifelines2. Each of the 318 patients is *aligned by* their 1^{st} occurrence of *IMC*, *ranked by* the number of *ICU* events, and *filtered by* the sequence of events.

research efforts have studied how end users interact with EHRs in hospitals. While some studies have focused on how patients can benefit from a display of their own EHR[24], most efforts have focused on how medical professionals as end users. These studies follow, for example, physicians' workflow in supplementing, annotating, and reusing EHRs [23, 26, 3, 18]. These field studies identify important design challenges in how EHR systems need to be improved to support medical professionals' tasks, unfortunately, they often fall short in recommending possible technologies for solving these problems [23, 3, 18].

Many EHR systems lack features that support important end-user tasks. Exploratory analysis, effective representation, and temporal queries are but a few that are often found lacking even in state-of-the-art systems such as Amalga[11] or i2b2[12]. As a result, many information visualization systems have been proposed different techniques to support these tasks and supplement the EHR systems. Some approaches are static visualizations, such as the one proposed by Powsner and Tufte[17], but most modern ones are interactive. Many of these support only a single EHR – Lifelines[15], Midgaard[1], Web-Based Interactive Visualization System[13] VIE-VISU[8], to name a few. They generally focus on supporting physicians to quickly absorb a patient's potentially lengthy medical history in order to make better medical decisions. On the other hand, a number of systems expand the coverage to multiple EHRs, for example, Similan[25], Protempa[16], Gravi++[7], VISITORS[10], and IPBC[4]. These systems typically focus on novel search and aggregation strategies for multiple EHRs.

These information visualization systems are all motivated by real issues physicians or clinical researchers encounter when the typical presentation of medical data is not conducive to their analysis tasks. However, because of limited availability of physicians and clinical researchers, very few systems have gone through multiple detailed long-term case studies. While these systems demonstrate the usefulness of their features in one or two isolated medical case

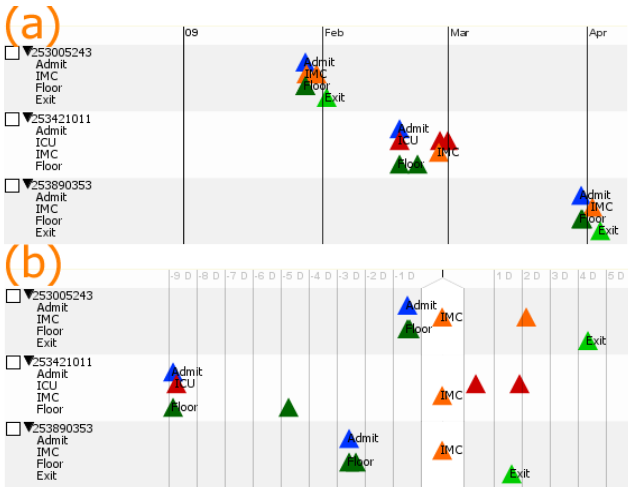

Figure 2: (a) shows three EHRs in Lifelines2 that are un-aligned (using calendar time line). (b) shows the same three EHRs aligned by their 1^{st} *IMC* event (time line is now relative).

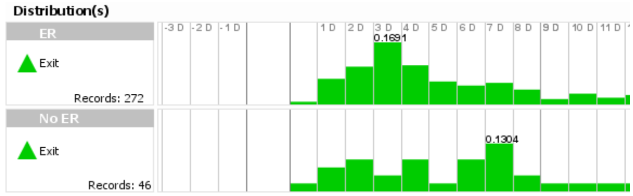

Figure 3: Comparison of the distribution of *Exit* events for two groups of patients. The top contains patients who go through ER, while the bottom one contains patients who do not.

studies, the results are harder to generalize. As a consequence, these information visualization efforts rarely make broader generalizations about their techniques. They also rarely make recommendations on the directions information visualization designers for EHRs should pursue further. In contrast, we applied Lifelines2 to eleven case studies, eight of which are medical scenarios according to the multidimensional in-depth long-term case studies (MILCS) model [19]. By analyzing the multidimensional user and usage data we collected, we believe we can contribute to the field by making some generalizations and recommendations. However, because Lifelines2 aims to support searching and exploring multiple EHRs, the generalizations and recommendations presented in this work may not apply to the design of single-EHR systems.

In addition to presenting the analysis on user and usage data of Lifelines2, we also present a process model which generalizes how physicians and clinical researchers seek information in EHRs. Our process model is similar in construction to the sense-making loop presented by Stuart Card and others [2, 14, 20]. However, ours differ in the level of granularity and application domain. We focus specifically on multiple EHRs and with a strong emphasis in temporal analysis. Our level of granularity and task-specificity is similar to the proposed process model for social network analysis [6].

3. LIFELINES2

Lifelines2 is designed for visualizing temporal categorical data for multiple records. Temporal categorical data are time-stamped data points that are not numerical in nature. For example, in an EHR, the patient's historical hospital visits, diagnoses, treatments, medication prescribed, medical tests performed, etc. can all be considered temporal categorical. These data are point data (no durations) with a name, and can be thought of as "events", This differs from temporal numerical data such as the results of medical tests, *e.g.*, blood pressure readings, or platelet counts over time. Lifelines2 visualizes these temporal categorical data by EHRs and provides a number of visualization and interaction techniques for analysts for exploratory analysis.

Figure 1 shows a screen shot of Lifelines2. (a) is Lifelines2's main display of ERH. Each patient occupies a row, and is identified by its ID to the left. Under the ID, a list of event types in that EHR is listed. Each event is represented by a color-coded triangle, which is placed on the time line. (c) is the control panel for Lifelines2. It indicates that each patient is aligned by the 1^{st} occurrence of *IMC*, ranked by the number of *ICU* events, and filtered by the sequence of events [*Admit*, *No ICU*, *IMC*, *ICU*]. EHRs that match the filter is highlighted in orange. Of the 318 EHRs in Figure 1, only 39 were found to be matches. (b) is a histogram temporally synchronized with (a). It is called a *temporal summary*, and it displays the distribution of *Admit*, *Exit*, and *ICU* events. In (a) and (b), analysts can zoom in, zoom out, pan, and scroll. Tool tips provide detailed information for each event when moused over.

By aligning all patients by its corresponding events (1^{st}, 2^{nd}, ect., and last, 2^{nd} to the last, etc.), physicians can better compare the patients as a group. Events that occur commonly before or after the alignment can be more easily detected. When an alignment is active, the time line becomes relative to the alignment (Figure 2 (b)). Analysts can also align by *all* occurrences of an event type, in which case Lifelines2 duplicates each EHR by the number of events of taht type the EHR contains, and shifts the individual duplicate by each of the event instances.

Analysts can rank the EHRs by their ID (default behavior), or by the number of occurrences of the different event types. They can also filter by the number of occurrences of event types, or by a sequence filter such as the one specified in Figure 1 (c). Align, Rank, Filter are affectionately called the ARF framework, and serves as a basis for user interaction in Lifelines2 [21].

Temporal summaries [22] are histograms of events over time, although analysts can change what it tabulates to number of EHRs, or number of events per EHR. Temporal summaries are temporally synchronized with the main visualization, and share the same temporal granularity. Analysts can use direct manipulation to temporally select the EHRs that contribute a certain bin in the histogram. By combining alignment and temporal summaries, analysts can select, for example, all patients that entered *ICU* within 24 hours of entering *IMC*.

Finally, after filtering and selection, analysts can choose, optionally, to save the results as a separate group. Multiple groups can be compared in comparison mode, where one temporal summary represents a group, and arbitrarily many groups can be compared. Figure 3 compares two groups of patients. The first contains patients who have entered emergency room, and the other contains those who have not. These patients are all aligned by their admission time (*Admit*), and the distribution of *Exit* events is plotted. The events are normalized by the number of patients in each group. This means that the bars represent the percentage of patients that exit in each day following their admission. There is a clear peak for patients who go through emergency room (ER), while those who do not have a much irregular distribution. The comparison features allow physicians to, for example, directly compare patient groups that undergo different treatment options to see which option produces better outcome.

4. A CASE STUDY

There are two phases in which we conducted our case studies using Lifelines2. In the first phase (early-adoption), physicians and hospital administrators work with us to iteratively refine and improve the features and usability of Lifelines2. They also learned the features of Lifelines2 during this time. This phase lasted for over a year, and we conducted three early-adoption case studies: (1) finding patients who exhibited contrast-induced nephropathy, (2) finding patients who exhibited heparin-induced thrombocytopenia, and (3) studying hematrocrit levels in trauma patients with respect to discharge patterns and length of stay in the hospital.

After the early-adoption case studies, we continually conducted eight additional mature-adoption case studies, five of which were in the medical domain. During the mature-adoption phase, no new features were implemented in Lifelines2. Only bug fixes and small features that facilitate the analysis were added. Using Lifelines2, our physician collaborators were able to (1) replicate the study that investigate the relationship between day light savings time change and heart attack incidents [9], (2) perform a follow-up study on heparin-induced thrombocytopenia in ICU patients [22], (3) study hospital room transfer in patients as a measure for quality assurance (two case studies, one for the Bounce-back patterns, and the other for the Step-up patterns), and (4) study the protocol change in when to use Bi-level Positive Airway Pressure (BiPap) and its impact on patient care.

We present below, in detail, a case study on patient room transfers to demonstrate how our physician collaborators typically apply Lifelines2's features in real scenarios. Hospital rooms can be roughly classified into the following types: (1) *ICU* intensive care units that typically provide the highest level of care) (2) IMC (intermediate medical care rooms that typically house patients who need elevated level of care, but not serious enough to be in *ICU* (3) *Floor* (normal hospital beds that typically house patients with no life-threatening conditions) (4) *Special* (emergency room, operating room, or other rooms). In this study, the dataset also include when the patients are admitted (*Admit*) and leaving the hospital (*Exit*) if they have already exited. In this study, each of the room data points comes with a time stamp, indicating when the patient is transferred-in. Transfer-out is implied by subsequent transfer-ins to another rooms or *Exit*.

The physicians are interested in the *Step-up* pattern. This is a pattern where a patient who was initially triaged to go to an IMC room, but then immediately escalated to higher-level care rooms. The pattern may be indicative of mistriage – that is, sending patients too sick to IMC instead of ICU in the first place. The exact criteria are patients who were sent to IMC, but escalated to ICU within 24 hours. For example, the fifth patient from the top in Figure 1 exhibits exactly the Step-up pattern.

There are two hypotheses our physician collaborators are interested in. First, the nurses in IMC have noticed anecdotally an increased number of step-up cases. Our collaborators want to verify the claim when compared with historical data. Secondly, our collaborators hypothesize that because newly graduated doctors enter the hospital in the third quarter (July-September) every year, the percentage of step-up cases may be higher in these months.

The original query seems simple: by aligning by all patients' *IMC* events, and selecting all *ICU* events that occur within 24 hours after the alignment, we should be able to identify all patients who exhibit the step-up case. However, when the authors and our collaborators examined data together, we realized several issues. For example, there should not be any *Floor* events between *IMC* and *ICU* (patient going from *IMC* to *Floor* then to *ICU*) because this suggests the escalation from *Floor* to *ICU* is likely not due to an earlier triage. Similarly, there should not be an *ICU* prior to the *IMC* in question. If there were, the patient was already in *ICU*, and this would be not be considered a Step-up. The visualization and the application of alignment made the discovery of these issues effortless. A direct application of, for example, SQL query using our original formulation would have missed these nuances. In Lifeliens2, these nuances can be handled and easily verified in Lifelines2 using the following procedure:

1. Perform a sequence filter using [*IMC*,*No Floor*,*ICU*], and save the results as a new group named `IMC-No Floor-ICU`.
2. Align by 1^{st} *IMC*.
3. Temporally select (in a temporal summary) *ICU* events that occur any time prior to the alignment, and remove the selected EHRs.
4. Temporally select *ICU* events that occur within 24 hours after alignment, and keep the selected EHRs.
5. Save as a new group and export this new group as a file.
6. Return to group `IMC-No Floor-ICU`.
7. Repeat steps 1-6 by changing the 1^{st} *IMC* to the n^{th} *IMC*. Stop when there are no records with n *IMCs*.

We conducted this study for every quarter from January, 2007 to December, 2009. Each quarter took about 12-20 minutes to performing. The data contains all patients who have been admitted to IMC in that period. A screen shot of a quarterly data is shown in Figure 1. Figure 4 shows

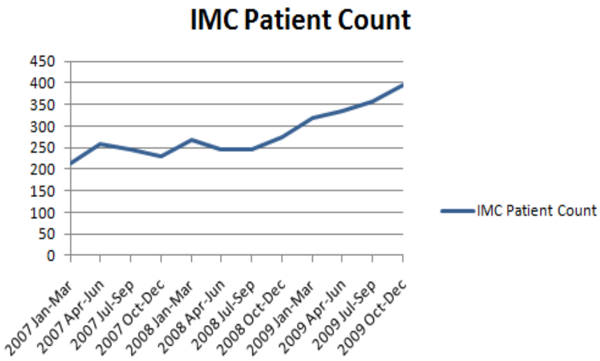

Figure 4: Number of patients admitted to IMC from the start of 2007 to the end of 2009.

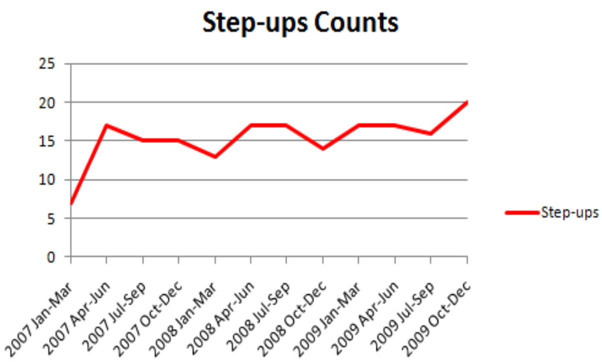

Figure 5: Number of patients who exhibited the Step-Up pattern from the start of 2007 to the end of 2009.

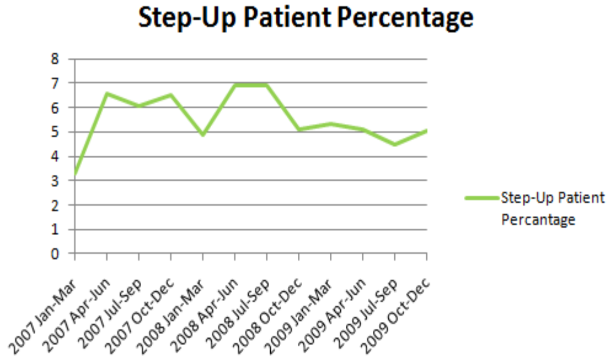

Figure 6: Percentage of patients who exhibited the Step-Up pattern from the start of 2007 to the end of 2009.

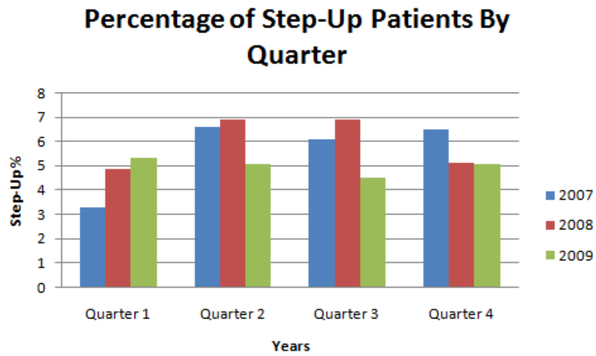

Figure 7: Percentage of patients who exhibited the Step-Up pattern from the start of 2007 to the end of 2009 by quarters.

the number of patients admitted to IMC during that period. It is clear that the IMC patient count is on a rising trend since the 3^{rd} quarter of 2008. Figure 5 shows the number of patients who exhibit the Step-up pattern in the same period, a subset of all patients admitted to IMC. The line graph is more jagged than Figure 4, however, its upward trending is also clear. Finally, we plot the percentage of patients who exhibit Step-up patterns (out of the IMC patient counts) in Figure 6. It shows that the percentage of Step-up patients peaks in the middle two quarters of 2008 (at nearly 7%), but has actually been in decline since then. This suggests that the anecdotal evidence the physicians received is, in fact, merely anecdotal, and does not represent a larger trend. However, these numbers also explained why the nurses feel otherwise – the number of total IMC patients have been steadily on the rise, and so the Step-up cases are also growing, reaching as many as 20 in the last quarter of 2009. One of our physician collaborators explains, "The nurses must have gotten the impression that mis-triaging occurred more often because they have encountered more Step-up cases. They felt the increased number of cases was due to errors in the triaging, while the real reason is more likely the increased of IMC patients." He also said, "The reason why there was an increase of IMC patients was not due to the increase of diseases or injuries. Instead, it was merely a reflection on the expansion of IMC care in the hospital."

To investigate the second hypothesis "in quarter 3 (Jul-Sep), because of the increase of new, less-experienced doctors, the number of Step-up cases would be higher.", the percentage of Step-up patients are ordered by quarter in Figure 7. The data does not lend support to the hypothesis. Of the four quarters, the second quarter has the highest number of Step-up cases altogether, and the first quarter has the fewest by a significant margin. There is no evidence of an increase of Step-up cases in quarter 3. One of our physician collaborator commented that, "The attending physicians (supervisors of the residents) must have been doing a good job reviewing the results of the resident triaging

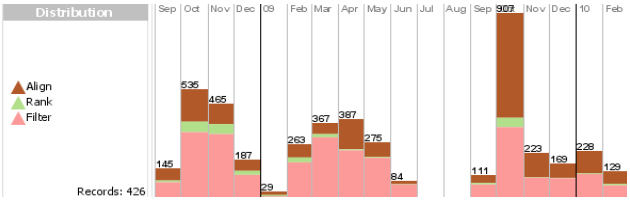

Figure 8: Distribution of *Align*, *Rank*, and *Filter* usage in Lifeliens2 over an year and half during which our case studies were conducted.

process." He, however, did not offer an explanation for why the numbers in the first quarter are so much lower than the others.

5. INTERACTION LOGS

The case studies such as the one presented in 4 demonstrate how Lifelines2 has been beneficial to analysts in a variety of different medical scenarios. However, some case studies use a set of features more than the others. For example, in replicating a study that links heart attack incidents to day light savings time change[9], analysts found the features in temporal summaries in conjunction with alignment are the most useful. On the other hand, in the Step-up study, a larger set of features (alignment, filter, temporal summaries, grouping and comparison) are all essential.

By September of 2008, most of Lifelines2 features were complete. Since then, Lifelines2 had been logging all the user actions. The logs keep track of analysts' interactions with Lifelines2 – features used, navigation actions, etc.. The Lifelines2 log output is in the format of Lifelines2 input, so the logs can be read by Lifelines2 for our analysis.

There were a total of 2477 logs. However, some of the logs were not due to exploring the data or demonstrating the exploration results to collaborators. Many of the logs were short, and no data files were opened aside from the default sample file. These are indicative that the session was due to testing and debugging uses. After removing the very short sessions, those that do not open at least one other file, and those that had logged no operators, 426 real sessions remain. We loaded these sessions into Lifelines2. The temporal summary in Figure 8 show the number of events of *Align*, *Rank*, and *Filter*. The minimal amount of activity in 2009 Jan represents winter break. The lack of any activity in July and August of 2009 represents the time during which the case studies were halted because the authors were out of town. The peculiar spike in October of 2009 represents an intensified analysis of the hospital transfer data with our physician collaborators. The amount of operations in that period was reflective of the fact that these case studies involved many steps, and over 15 different datasets.

Table 1 (Page) summarizes the logs of the usage of Lifelines2 for the 426 sessions. The operations are broken down into the following five main categories: ARF, Temporal Summary, Comparison, Data Operations, and Navigation. The table includes the raw number of counts, counts per session, and percentage of sessions that logged such operations.

Operation	Count	Ave.	%
ARF			
Align	1680	3.94	87%
Rank	260	0.610	42%
Filter	2564	6.019	68%
Temporal Summary			
Show Summary	623	1.46	30%
Temporal Selection	531	1.25	24%
Comparison			
Event Type Change	406	0.95	11%
Comparison Type Change	79	0.18	12%
Group Change	406	0.95	12%
Distribution Type Change	179	0.42	13%
Data Operations			
Keep Selected	400	0.94	30%
Remove Selected	96	0.23	18%
Save Group	409	0.96	29%
Change Group	687	1.61	23%
Navigation			
Zoom In	646	1.52	30%
Zoom Out	157	0.37	13%
Time Range Slider	1865	4.38	29%
Change Granularity	217	0.51	14%
Scroll	6840	16.1	100%
Collapse	55	0.13	8%
Expand	30	0.07	5%

Table 1: Operator usage in Lifelines2: Raw counts, average count per session, and percentage of sessions in which the feature is used.

With respect to the ARF Framework, *Filter*was the most-frequently used operator. *Alignment* was second, and trailed by *Rank*. However, a larger percentage (87%) of sessions recorded at least one use of *Align*, while only 68% had any *Filter*. While *Rank* was useful to reorder the records by their event counts, it was ultimately not a vital operator in our case studies. When pairs of *Align*, *Rank*, and *Filter* were looked at as sequences, [*Align*,*Filter*] and [*Filter* , *Align*] occurred in 250 (59%) and 211 (50%) sessions respectively. [*Align*, *Rank*] and [*Rank*, *Align*] occurred at 162 (38%) and 123 (29%) sessions respectively. Finally, [*Rank*, *Filter*] occurred in 148 (37%) sessions, and [*Filter*, *Rank*] occurred in only 48 (11%) sessions. When looking at sequences of three operators, the break down (number of sessions that had the contained the sequence) is as follows: ARF ([*Align*, *Rank*, *Filter*]): 123, AFR: 41, FAR: 37, FRA: 27, RFA: 104, and RAF: 87. These numbers indicate that although *Rank* is the least popular of the three operations, when it is used, *Rank* is typically used prior to *Align* or *Filter*, or both. On the other hand, *Align* and *Filter* are used fairly frequently, and which one tends to be used before the other depends on the analysts' tasks.

30% of the sessions used temporal summaries, and 24% used selections in temporal summary. The average number of these operations across all sessions were over 1 per session. This means that in sessions that these operations were used, they were used many times, so much that the average count per record is brought up.

The operations under the Comparison feature only occurred in 11-13% of all sessions. However, analysts tended to change the event types in the comparison and the groups in the comparison heavily. Changing the type of comparison (Between Group/Within Group/Both) or the object of aggregation (Events/Reocrds/Events Normalized By Record Count) were less frequently used.

In these EHR case studies, analysts tended to use *Keep Selected* as opposed to *Remove Selected* in conjunction with filtering. *Save Group* occurred in 29% of the sessions while *Change Group* occurred only in 23%. This indicates that for some datasets, analysts would save a group, but not change into that group specifically. This situation occurs not because the analysts do not look at the newly created group. Instead, this is because Lifelines2 automatically brings the analysts to the newly created group without having them perform the group change themselves. By raw counts, *Change Group*, as expected, is used more frequently than *Save Group*.

The first thing to notice in the navigation operations is that *Scroll* (to pan vertically) is a dominant operation. Every session involved scrolling, and on average, each session has more than 16 scrolls. Each scroll is captured when analysts let go of the mouse on the scroll bar or the up/down, page up/page down keys. Changing the *Time Range Slider* (to zoom or pan horizontally) was a distant second in usage in this category. In contrast, *Change Granularity* (temporal granularity) was not as popular. This may be attributed to the fact that Using the *Time Range Slider* analyst can control the temporal range more finely. Even after using the cruder *Change Granularity*, an adjustment in the *Time Range Slider* was often necessary. *Zoom In* was used more often than *Zoom Out*. This is attributed that users can perform zoom out by using *Change Granularity* or use the *Time Range Slider*. *Collapse Record* and *Expand Record* were the least used features. These features collapses the vertical space of each EHRs to that more can fit in one screen, or expand them to see details more clearly.

6. A PROCESS MODEL FOR EXPLORING TEMPORAL CATEGORICAL RECORDS

Combining the log data, observations with the collaborators, and interviews and comments with our physician collaborators, we constructed the following process model for exploring temporal categorical records.

1. Data Acquisition
2. Examine data in visualization for confidence (overview/browse)
3. Exploratory Search
   (a) Iteratively applying visual operators
   (b) Evaluate Results of manipulation
   (c) Deal with unexpected findings
4. Analysis, Explanation
   (a) Examine path of search as a whole
   (b) Determine to what extent the questions are answered
      i. At the limitation of the system
      ii. At the limitation of the data
   (c) Refine existing questions
5. Report results to colleagues
   (a) Document discovery
   (b) Disseminate data
6. Move onto new questions

The information seeking process begins with data acquisition. Since Lifelines2 is not directly linked to the numerous databases a hospital may have, we obtain our data through our physician collaborators. At the beginning of a medical case study, the analysts decide the scope of the data that they want to examine. We then request database administrators of the hospital to gather the required data via SQL. We then deidentify and preprocess them for our case studies. During the information seeking process, sometimes the analysts return to the data acquisition stage because they (1) become unsatisfied with the data, (2) found systematic errors in the data, or (3) want to incorporate more data for deeper analyses.

In our case studies, this stage typically takes a long time. The reason for the lengthiness in acquiring data lies in the complexity of the data, infrastructural or organizational barriers. For example, the desired data may reside in different databases, named using different IDs or codes, and the lack of documentation on the mapping from medical terminology to the terms used in the database schema. They may be difficult to search, and may require first finding someone who knows where it is. This stage can take from days to weeks, depending on how involved the case study is, ease in locating data, and availability of physician collaborators and database administrators.

When de-identification and preprocessing are complete, our physician collaborators would examine the data visually. They would cursorily browse and sometimes examine in detail the data to make sure the data reflects what they know. One of the most common and immediate results in using a visualization system to examine data for the first time is the discovery of interesting artifacts (systematic errors, lack of consistency, etc.). For example, when the data in the mature heparin-induced thrombocytopenia case study was first converted, our physician collaborators found that some patients were given drugs *after* they had been discharged dead! In this particular case, the discovery was made by playing with the Lifelines2 interface by showing the distribution of drug prescriptions and aligning by discharged dead events. The summary showed suspicious timing, and when we tracked the patients down, we found 7 of them were prescribed drugs after they had been discharged dead. All 7 cases occurred within one hour of their discharge dead events, and our physician collaborators concluded that this occurred because the drug database must have recorded the prescription event with some delay. For that case study, this incident raised questions on how reliable the time stamp was, and whether subsequent case study would be affected. We eventually found better data to circumvent this particular systematic problem. Sometimes, however, the data is not usable or is discovered to be unsuitable, so we would take a step back to the data acquisition stage.

After analysts have gained confidence in the data and the visualization (that the visualization was not giving false impressions somehow or missing data), they move onto Stage 3 – Exploratory Search. They would start seeking answers to their preconceived questions or finding evidence for their hypothesis. However, in the process of seeking answers to one question, new questions often spawn when they notice interesting or unexpected data. At this point they would utilize their domain knowledge to try to explain what they see (for example, narrate about certain EHRs to aid their reasoning), or they would write down new questions for later exploration (for example, noticing that a characteristic of the event pattern and needing a way to handle them).

Analysts approach exploratory search differently. We have observed that analysts would apply alignment on different sentinel events in the same exploratory session to look at the data in different views. By using different alignment while showing distribution of certain events they care about, they aimed to find useful or telling "sentinel" events. Some analysts used a more traditional way for exploration: actively manipulate the display by ranking, filtering iteratively, or changing the temporal summary. Regardless of the strategy they used, alignment remained the strongest indicator on their focus on data. A change in the alignment event indicates a change of exploratory focus. Sometimes when the collaborators had aligned by one event, and realized that alignment would not lead them to the information they wanted to see, they would reformulate the question and subsequently used a different alignment. This had been observed with multiple physicians.

Another important observation of the analysts in exploratory search was that the analysts paid special attention to the change of data at each step of the interaction. The Lifelines2 log data shows how often the physicians used *Scroll* to view records in the dataset. Closer analysis revealed that there are two hot spots where many scroll operations are performed. The first is when the analysts were examining the data in Stage 2. The second is after each align, rank, or filter operator had been applied. For example, identifying the nuances of the Step-up query was accomplished in this step.

Aside from manual scrolling, our collaborators would also keep an eye on (1) the distribution of their favorite events in the temporal summary and (2) the number of records in view. We observed that when align, rank, filter and group operations are applied, the analysts would focus on these two things. They give the analysts a global feel of how the data is changing when they apply a variety of operations. In fact, in cases where heavy exploration was required, as in the early adoption heparin-induced thrombocytopenia case study, we noticed that the physicians kept their eyes fixed on the temporal summary as a variety of filters were applied. When we worked on the mature heparin-induced thrombocytopenia study [22], different physicians also showed the same tendency to focus on the temporal summaries as the data is being sliced and diced. By fixing their attention on the temporal summaries, they could get a good sense of what changed in the data, and how the operations they had chosen to perform changed it. They could then decide if they were on the right path of exploration. If they did not like a previously applied filter, they would backtrack to that previous state, and rethink their approach. In fact, our collaborators would even use the comparison feature on several previously created groups to examine if the paths of exploration seemed to be a fruitful one. Temporal summaries became an indispensable mini map to the thousands of patient records that cannot fit on the screen. Although focusing on temporal summaries was quick, our collaborators would still examine the records individually when they had the chance, though not exhaustively.

When our collaborators encountered unexpected discoveries, they would save the current data (with the discoveries) as a new group, and made a mental note to themselves to come visit the saved group. When they wanted to continue exploration using the data at a later time, they would export the current data into a new file. When they found something noteworthy, the annotation tool was used, sometimes in conjunction with the built-in screen capture in Lifelines2 – although annotation was recorded to only occur in 5% of all logged sessions (not listed in Table 1.

When our collaborators arrived at an interesting point where their questions might be answered, they would use their domain knowledge to analyze what they saw, and offer explanations (Stage 4: Analysis, Explanation). They would verify how they got there by looking at the groups they created before, or by examining them all at once in the comparison view, as if they were double-checking their work. When they were satisfied with what they saw, they would examine the data, and decide whether their questions could be, or were, answered.

Sometimes a dead-end was reached, and they would realize that more data was required, leading back to Stage 1. However, sometimes the dead-end was encountered because we were at the limitations of Lifelines2. This occurred when the analysis in a case study required features Lifelines2 simply does not support (*e.g.*, numerical values, percentage changes). In these cases, the exploration would come to a halt. Unless we had found a solution (a good way to bin the numerical values into categorical values, for example), the case study would discontinue. Sometimes the limitations of Lifelines2 could be compensated by other systems. For example, in the heart attack and daylight savings case study, we resorted to using Excel as a platform to compute average incidents per day. In these cases, the case study may come to a stop but with fruitful results. Through this exploration process, our collaborators sometimes found that their original questions could not be answered or was not suitable. They would refine their questions. If they had noticed something during the exploration, they might choose to pursue the question with a different spin. The value of the Lifelines2, at that point, would be allowing the analysts to discover the questions they did not have before.

Finally, when a case study is completed and analysts arrive at a satisfactory point, analysts would prepare their findings. Our collaborators routinely keep subsets of EHRs that represent the fruit of their labor, screen shots, annotations, and spreadsheets created in our collaborative session. They would take these data to show their colleagues or supervisors to argue for or against a procedure/policy change. They would also argue for the usage of technologies such as Lifeines2 for their daily work. From the perspective of presentation, Lifelines2 visualization seems to be adequate. We had not heard requests from our collaborators that screen shots needed to be modified for other non-collaborator physicians' consumption. Our collaborators are capable of explaining what is happening in a Lifelines2 screen shot.

7. RECOMMENDATIONS

The case studies, Lifelines2 logs, and observations have revealed some interesting user behaviors when dealing with multiple EHRs. They have also revealed the strengths and weaknesses of Lifelines2. We generalize these into the following six design recommendations for future developers of visualization tools for multiple EHRs.

1. **Use Alignment** The usefulness of alignment was evident in the Lifelines2 logs, and from observations and collaborator comments. The user logs corroborate the findings of alignment in our previous controlled experiment[21]. When dealing with a large number of EHRs, the ability to use alignment to impose a strict relative time frame was important to our collaborators. It allowed for quicker visual scanning of the data along the alignment. The dynamism of alignment allowed the analysts to quickly switch perspectives and focus if they need to. The idea of "anchoring" the data by data characteristics for exploration had been successful in other visualization systems, and alignment seems to be one natural version of it for the temporal domain. Developing future visualization systems for EHRs should leverage on alignment for its power, flexibility, and wide range of applicability. However, there may be other "anchoring" techniques in temporal visualization in perhaps different situations, and it is worth pursuing them.

2. **Show Details** One surprise finding was that our collaborators liked to look at the details of the records. One piece of evidence is that *Scroll* was the most frequently-used operation. It is indicative of how much the analysts liked to view records in detail. Seeing the details and being able to compare the detailed records that are close to one another seem to reassure the analysts that no data are missing, broken, or lost along the visualization pipe or the analysis process. Another piece of evidence was that the *Collapse* operator, which makes details harder to see, was hardly used. Our second recommendation is that detailed depiction of the records is important. Even if the primary view of the data was to be in an overview display, the analysis tool must always make the details accessible to analysts – and preferably for many records at once.

3. **Overview Differently** Lifelines2 provides an overview in the form of temporal summaries. While it is perhaps the most important feature after ARF (and one of the most frequently used features), more ways to overview the data may be beneficial to analysts. We observed that during iterative filtering , our collaborators tended to focus on the overview most of the time to get a sense of what each filtering operator does, and then examine some records in detail. Other types of overview can complement temporal summaries. For example, in additional the "horizontal" temporal summaries, "vertical" overviews can simultaneously show a different aggregation over records. Furthermore, a good vertical overview design may reduce the amount of *Scroll* necessary.

4. **Support Richer Exploration Process** The features in Lifelines2 that support branching in exploration are *Save Group* and *Change Group*. The rudimentary features were both used fairly frequently (averaging almost one use per session in the log), and a lot of improvements are desired. As analysis process becomes more and more involved, information visualization systems need to better support branching searches, history keeping, and backtracking. Another important feature is to allow users to perform set operations on groups of EHRs. Union, intersection, and difference are very common operations, and can prevent many instances where users have to switch to a different tool in order to perform them.

5. **Flexible Data Type** Some of the earlier case studies stopped because Lifelines2 does not provide support to numerical values. Numerical values needed to be first converted into categories (*e.g.*, *High blood pressure*, *Normal blood presser*, etc.). While sometimes the categories suffice, numerical values are important as well. Working with our collaborators, we discovered that depending on the focus of a medical scenario, sometimes our collaborators reasoned at a coarser granularity (categories), and sometimes finer (numerical values), and sometimes the granularities change within the same scenario. Most visualization systems focus on either categorical data or numerical data. It would be an interesting direction to investigate how use different granularities to support human reasoning, as opposed to using them for machine consumption [10, 16].

6. **Higher Information Density** The amount of scrolling we recorded indicates that the amount of data our collaborators want to see is typically much larger than a screen can hold. It would be important to improve information density in Lifelines2, and other time-line based visualizations, *e.g.*, [25, 1, 5]. It is worth mentioning that a good solution to better "vertical" overview may solve this problem at the same time. The designers need to focus on the specific tasks end-users wish to perform.

8. CONCLUSIONS

We present a generalization of our eight case studies with EHRs using Lifelines2. By analyzing the feature usage data, user comments, and study observations, we present an information seeking process model for multiple EHRs and a list of recommendations for future information visualization designers for EHR systems. While some of our results are limited to capabilities of Lifelines applied to EHRs, we were able to draw some more general recommendations. We encourage the information visualization field to continue building a user-centered, task-based design requirements and process model for the betterment of EHR end-users.

9. ACKNOWLEDGMENTS

We would like to thank Dr. Mark Smith, Dr. Phuong Ho, Mr. David Roseman, Dr. Greg Marchand, and Dr. Vikramjit Mukherjee for their collaboration. We would like to thank MedStar One for their continued support of our work. We would also like to thank Krist Wongsuphasawat for his help in the case studies.

10. REFERENCES

[1] R. Bade, S. Schelchtweg, and S. Miksch. Connecting time-oriented data and information to a coherent interactive visualization. *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 105–112, 2004.

[2] S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers, San Francisco, CA, USA, 1999.

[3] Y. Chen. Documenting transitional information in emr. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 1787–1796, New York, NY, USA, 2010. ACM.

[4] L. Chittaro, C. Combi, and G. Trapasso. Data minding on temporal data: a visual approach and its clinical application to hemodialysis. *Journal of visual Languages and Computing*, 14:591–620, 2003.

[5] DataMontage. http://www.stottlerhenke.com/datamontage/.

[6] D. L. Hansen, D. Rotman, E. Bonsignore, N. Milić-Frayling, E. M. Rodrigues, M. Smith, and B. Shneiderman. Idd: An interactive distributed debugger. Technical Report CU-CS-274-84, Department of Computer Science, University of Colorado at Boulder, 1984.

[7] K. Hinum, S. Miksch, W. Aigner, S. Ohmann, C. Popow, M. Pohl, and M. Rester. Gravi++: Interactive information visualization to explore highly structured temporal data. *Journal of Universal Computer Science (J. UCS) – Special Issue on Visual Data Mining*, 11(11):1792–1805, 2005.

[8] W. Horn, C. Popow, and L. Unterasinger. Support for fast comprehension of icu data: Visualization using metaphor graphics. *Methods of Information in Medecine*, 40(5):421–424, 2001.

[9] I. Janszky and R. Ljung. Shifts to and from daylight saving time and incidence of myocardial infarction. *New England Journal of Medicine*, 359(18):1966–1968, 2008.

[10] D. Klimov, Y. Shahar, and M. Taieb-Maimon. Intelligent selection and retrieval of multiple time-oriented records. *Journal of Intelligent Information Systems (Published Online)*, 2009.

[11] Microsoft. http://www.microsoft.com/amalga/.

[12] S. Murphy, M. Mendis, K. Hackett, R. Kuttan, W. Pan, L. Phillips, V. Gainer, D. Berkowicz, J. Glaser, I. Kohane, and H. Chueh. Architecture of the open-source clinical research chart from informatics for integrating biology and the bedside. In *Proceedings of the American Medical Informatics Association Annual Symposium (AMIA '07)*, pages 548–552, 2007.

[13] D. S. Pieczkiewicz, S. M. Finkelstein, and M. I. Hertz. Design and evaluation of a web-based interactive visualization system for lung transplant home monitoring data. *Proceedings of the American Medical Informatics Association Annual Symposium (AMIA '07)*, pages 598–602, 2007.

[14] P. Pirolli and S. K. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of the 2005 International Conference on Intelligence Analysis*, page 6, 2005.

[15] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman. Lifelines: Visualizing personal histories. *Proceedings of CHI 96*, 1996.

[16] A. R. Post and J. H. Harrison. Protempa: A method for specifying and identifying temporal sequences in retrospective data for patient selection. *JAMIA*, 2007.

[17] S. Powsner and E. Tufte. Graphical summary of patient status. *The Lancet*, 344:386–389, 1994.

[18] A. Sarcevic. "who's scribing?": documenting patient encounter during trauma resuscitation. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 1899–1908, New York, NY, USA, 2010. ACM.

[19] B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. *Proceedings of BELIV '06*, pages 38–43, 2006.

[20] J. J. Thomas and K. A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center, 2004.

[21] T. D. Wang, C. Plaisant, A. J. Quinn, R. Stanchak, S. Murphy, and B. Shneiderman. Aligning temporal data by sentinel events: discovering patterns in electronic health records. *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, pages 457–466, 2008.

[22] T. D. Wang, C. Plaisant, B. Shneiderman, N. Spring, D. Roseman, G. Marchand, V. Mukherjee, and M. Smith. Temporal summaries: Supporting temporal categorical searching, aggregation, and comparison. *IEEE Transaction on Visualization and Computer Graphics*, 15(6):1049–1056, 2009.

[23] L. Wilcox, J. Lu, J. Lai, S. Feiner, and D. Jordan. Physician-driven management of patient progress notes in an intensive care unit. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 1879–1888, New York, NY, USA, 2010. ACM.

[24] L. Wilcox, D. Morris, D. Tan, and J. Gatewood. Designing patient-centric information displays for hospitals. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 2123–2132, New York, NY, USA, 2010. ACM.

[25] K. Wongsuphasawat and B. Shneiderman. Finding comparable temporal categorical records: A similarity measure with an interactive visualization. *Proceedings of IEEE Symposium on Visual Analytics Science and Technology (VAST '09)*, pages 27–34, 2009.

[26] X. Zhou, M. S. Ackerman, and K. Zheng. Doctors and psychosocial information: records and reuse in inpatient care. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 1767–1776, New York, NY, USA, 2010. ACM.

APPENDIX

A. PREFERRED REVIEW APPROACH

- **Primary focus**: Computing (information visualization, HCI)
- **Secondary focus**: Usage of Electronic Health Records

3 topics covered in the paper:

- Display and visualization of medical data

- Human-centered design of health informatics systems
- User-interface design issues applied to medical devices and systems