Published in final edited form as:

Comput Biol Med. 2015 April 1; 59: 221–231. doi:10.1016/j.combiomed.2013.12.006.

Clinicians' Evaluation of Computer-Assisted Medication Summarization of Electronic Medical Records

Xinxin Zhu^{a,*} and James J. Cimin^b

^aIBM T.J. Watson Research Center, U.S.A

^bNational Institutes of Health, U.S.A

Abstract

Each year thousands of patients die of avoidable medication errors. When a patient is admitted to, transferred within, or discharged from a clinical facility, clinicians should review previous medication orders, current orders and future plans for care, and reconcile differences if there are any. If medication reconciliation is not accurate and systematic, medication errors such as omissions, duplications, dosing errors, or drug interactions may occur and cause harm. Computer-assisted medication applications showed promise as an intervention to reduce medication summarization inaccuracies and thus avoidable medication errors.

In this study, a computer-assisted medication summarization application, designed to abstract and represent multi-source time-oriented medication data, was introduced to assist clinicians with their medication reconciliation processes. An evaluation study was carried out to assess clinical usefulness and analyze potential impact of such application. Both quantitative and qualitative methods were applied to measure clinicians' performance efficiency and inaccuracy in medication summarization process with and without the intervention of computer-assisted medication application. Clinicians' feedback indicated the feasibility of integrating such a medication summarization tool into clinical practice workflow as a complementary addition to existing electronic health record systems. The result of the study showed potential to improve efficiency and reduce inaccuracy in clinician performance of medication summarization, which could in turn improve care efficiency, quality of care, and patient safety.

Keywords

Electronic health record; Medication error; Medication summarization; Timeline; Performance efficiency; Performance inaccuracy

© 2013 Elsevier Ltd. All rights reserved.

*Corresponding author. Tel.: +1914 945 4916. zhux@us.ibm.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Data from hundreds of organizations show that inaccurate communication of medical information at patient transition points is responsible for as many as 50 percent of all medication errors and up to 20 percent of adverse drug events in hospitals (Barnsteiner, 2005). “Medication reconciliation” is the process of resolving discrepancies and supporting medication ordering by summarizing an accurate list of a patient’s medications (IHI, 2009). Failure in reconciling medications accurately and systematically could cause medication errors such as omissions, duplications, wrong dosage, or drug interactions. However, clinicians must deal with large volumes of clinical data in their workflow and make critical clinical decisions quickly, often under time pressures, which limit exhaustive review of all details of a patient’s medication history from multiple data sources. Gathering medication information over time from a patient record is a labor-intensive process that involves addressing multiple information needs.

Given clinicians’ time limitations, an automated system that accurately presents medication information extracted from different data sources over time can help clinicians get a quick, dynamic, and reliable summary of patient medication information. An application such as this not only can help support care providers in gathering the needed medication information under time pressure, but also has much potential in reducing errors and improving care efficiency, clinical decision making, quality of care, and patient safety. We hypothesized that a medication timeline that captured critical medication information could help clinicians organize and manage patients’ medications in a reliable and consistent way. This study evaluated the clinical usefulness of a previously developed prototype computer application (Zhu et al., 2009) that can visualize medication information in timelines for clinicians, using open source software called (Timeline, 2006–2009). Medications were parsed and mapped to the event taxonomy, and then represented in Timelines.

The study subjects were six Columbia University Medical Center (CUMC) clinicians who had full access to Web-based Clinical Information System (WebCIS) (Hripcsak et al., 1999), a CUMC clinical results review and documentation system and consisted of interviews, observations, and questionnaires ranging from 60 to 90 minutes. WebCIS provided narrative medication lists as part of clinic notes, admission notes, and discharge summaries along with coded medications from the outpatient medication order entry system and the inpatient pharmacy system.

Six real patients’ records were used for this study. Interviews mainly assessed subjects’ experience in clinical practice and with electronic health record (EHR) systems. The observation sessions assessed and compared how clinicians accessed, assembled and drew inferences from multi-source medication information in a lengthy patient record with and without help from a developed medication summarization Timeline application. Questionnaires surveyed subjects’ opinions on the usefulness of the medication Timeline application. Data collected during the sessions were used to evaluate: 1) the application’s usefulness, 2) how well the software addressed clinicians’ information needs on medications, and 3) if the medication reconciliation Timeline application could help clinicians with their medication information gathering.

The study results validated that using Timeline to represent medication data extracted from various sources would assist clinicians in reducing inaccuracies in the medication summary process when gathering medication information from electronic health records. In certain cases, one can potentially reduce the percentage of inaccurate inpatient medications by more than 50% with the help of Timeline. Findings from this study can also be used to improve on the design and development of future automated tools that summarize medication information for clinicians.

2. Methods

2.1. Subjects

Six medical residents from the Department of Medicine at NewYork-Presbyterian Hospital (NYPH) who had access to WebCIS were recruited to be study subjects.

2.2. Setting

For each subject, there was an interview plus observation session on how they use WebCIS and Timeline to perform a list of clinical tasks. All interviews with subjects were conducted and finished within two consecutive weeks. There was no change in WebCIS or Timeline features during that period.

The interviews took place at the Columbia University Department of Biomedical Informatics. A video capture and analysis tool, (Morae, 1995–2009), was used to record a video of all computer screen activities and the voice of the interviewed users. These features helped capture study subjects' interactions with the systems and enabled careful scrutiny of each subject's answers and performance, noting evident and potential usability problems.

2.3. Selection of Clinical Scenarios

This study randomly chose six clinical scenarios from records of deceased patients who were treated at CUMC to avoid dynamic changes in patient records after the completion of the gold standard reference list.

In this study, Medical Record Numbers of deceased patients presented the following clinical scenarios:

Scenario 1: 78-year-old woman with multiple co-morbidities including type II diabetes (DM-II), hypertension (HTN), advanced dementia (bed-bound, minimally verbal), Stage V chronic kidney disease (CKD), untreated colon cancer, iron deficiency anemia (IDA), arthritis both knees, congestive heart failure (CHF), renal osteodystrophy, hyperlipidemia and depression was recently admitted to hospital.

Scenario 2: 51-year-old woman with past medical history of mild right-side weakness, esophagitis, hiatal hernia, mild prior dysphagia, prior smoking, childhood asthma without history of intubation reported sudden onset of right then left hip pressure pain, radiating to mid femur.

Scenario 3: 80-year-old female with past history of acute renal failure secondary to malnutrition, dehydration, rhabdomyolysis, renal stones, mild dementia with anxiety and depression was recently admitted.

Scenario 4: 86-year-old woman with hypertension, end stage renal disease (ESRD) (on peritoneal dialysis since 2007), with episodes of spontaneous bacterial peritonitis (positive Methicillin-resistant *Staphylococcus Aureus*), intracerebral hemorrhage (ICH) on anticoagulation, presented to emergency department with weakness, altered mental status.

Scenario 5: 86-year-old woman with history of hypertension, poorly controlled type II diabetes on insulin, renal insufficiency, peripheral vascular disease (PVD), hypercholesterolemia, left bundle branch block (LBBB) (present since at least 1999), who presented with worsening exertional dyspnea and acute onset of chest pain.

Scenario 6: 87-year-old gentleman with history of hypertension, hyperlipidemia who presented with productive cough (dark brown sputum) for a week, dyspnea, overall malaise, weakness and decreased appetite. Patient also noted fevers and rhinorrhea, denied chest pain nausea, vomiting and diarrhea.

2.4. Intervention —Medication Timeline Application

The application was designed according to analyses of a cognitive study, to be published in a separate paper, of clinicians' thought processes in medication summarization and developed through multiple iterations with clinicians. A prototype of a temporal visualization system was implemented using an open source software package called Timeline (Zhu et al., 2009). Specific methods were applied to address research issues in the areas of temporal and clinical uncertainties, temporal and clinical granularities, drug class discrepancies, and graphic representation and visualization of multi-dimensional medication data. Medications were parsed and then represented in Timelines in different shapes and colors based on their duration, temporal events, and clinical certainty of patient usage. Each medication entry was linked to a pop-up window, which includes its data source and the portion of the electronic patient record where the medication was mentioned. In addition, users can filter and highlight certain medications based on different data sources. Figure 1 illustrates an example using Timeline software features, where a user can highlight discharge and inpatient medications from the complete medication list. In this illustration, to find out what the discharge medications are, blue is selected as the highlighting color. As "discharge" and "inpatient" were typed in the text box, the highlight function is automatically triggered, and all discharged medications are highlighted with blue squares, and inpatient medications with red.

2.5. Gold Standard

The gold-standard medication information for this evaluation was generated as shown in Figure 2: two experts read patient records individually and provided medication information to the clinical questions in Table 1 for each patient. When applicable, the following medication attributes were collected in those answers regarding medication orders: medication name, dose, route, frequency, data source, context, prescription time, start time, stop time, and usage status. For each record, medication information on which they

disagreed was collated and built into a single document by a moderator and sent back to reviewers for further consideration. A reviewer could comment on the disagreed medication information, agree or reject the other's opinions, or change his or her responses to the other's opinions until all the disagreements were resolved. In the end, they reached agreement on all the medication information, and the gold standard reference list was formed based on their consensus.

2.6. Study Procedure

The study was designed to last up to 90 minutes. First, the clinicians were asked to respond to a series of questions in a brief semi-structured, recorded interview. Following is a list of questions asked:

1. Please describe your experience with computers and EHRs such as WebCIS.
2. Please tell us how long you have been working as a resident at NYPH.
3. Please list challenges you have experienced in attempting to understand a patient's medication history.

The intervention of this study, a system-generated medication summary in timelines (also known as the medication Timeline application, abv. Med. Timeline), was introduced to all subjects in an orientation session. During the orientation, a sample patient record was used to show subjects what functions the medication Timeline application supported and how to browse and interact with Timeline. Subjects could spend 5–10 minutes to get familiar with the user interface and Timeline functions.

In the observation sessions, six scenarios from deceased patients were presented to study subjects. Those patients had been hospitalized in CUMC before they deceased, and their records included discharge summaries, inpatient and outpatient visits, including pharmacy orders from Eclipsys XA. Subjects were seated in front of a computer and asked to complete a series of tasks using WebCIS only or WebCIS plus Timeline. Two types of patient records (Type I: original patient records and Type II: records with additional information prepared by the medication Timeline applications) were mixed before assignment to study subjects. Patient records were shuffled to reduce potential selection bias and contaminating factors across subjects and records; thus, individual subjects viewed different combinations of records. Study subjects receiving records with additional information provided by the Timeline application could refer back and forth to the original records when reviewing the system-generated medication Timeline views. The study subjects were presented with patient records according to the assignment table (Table 2).

For each assigned record, subjects were asked to perform the tasks shown in Table 1, and the time spent by each subject on each task for each record was recorded. They were asked to write down their answers in Word documents, which were then collected and compared to the “gold-standard” medication information to measure their performance accuracy. These five tasks were derived from the cognitive study, where key process components, data sources, and medication attributes were identified to be critical for clinicians' mental process of medication summarization.

At the end of the study sessions, subjects were asked to answer a brief survey (5–10 minutes). The survey questionnaire (see Appendix for full survey), which was intended to assess the usefulness of the proposed medication timeline application, consisted of twelve questions followed by a free-text comment field, where study participants were encouraged to write down their suggestions and opinions about the study. Sample survey questions are shown as in Table 3.

2.7. Data Analysis

2.7.1. Quantitative Analysis—There were two major areas for quantitative analysis: first, was performance efficiency measured as time spent on finishing tasks; second, was performance inaccuracy. Time spent was useful for analyzing how the Timeline application could potentially improve clinicians' efficiency, and performance inaccuracy was valuable to evaluate the effectiveness of how the proposed application could help clinicians reduce medication errors. Descriptive analyses combined with statistical measures in each area were described in the following sections.

2.7.1.1. Performance Efficiency: In this study, clinicians' performance efficiency was measure by time. As shown in Table 2, six cases were reviewed under two system circumstances, one being WebCIS, and the other being WebCIS plus Timeline. In total, there were twelve different case-system combinations, and each case-system combination was accomplished by three different subjects. The average time spent by subjects on each case under the two system circumstances were calculated and compared. Because each case had five tasks, the average time spent by subjects on each task under the two system circumstances was calculated and compared. Comparisons were made between the following pairs under two system circumstances:

1. Average time spent on all tasks.
2. Average time spent on task to find out admission time.
3. Average time spent on task to find out inpatient medications.
4. Average time spent on task to find out medication history.
5. Average time spent on task to find out discharge medications.
6. Average time spent on task to find out medication classes.

2.7.1.2. Performance Inaccuracy: In this study, clinicians' performance inaccuracy was measured by the inaccuracy rate, which is a percentage calculated as the number of inaccurate medications found by clinicians divided by the total number of medications in patient records. It is worth mentioning that the reason why inaccuracy instead of accuracy was deliberately chosen here is because granular categories of inaccuracy can be further analyzed to shed more lights on causes of medication errors for future studies.

As seen in Table 1, for task 1, subjects were asked to write down a time for each patient's admission time, and their answers were graded as *correct* or *incorrect*.

For tasks 2–4, subjects were asked to provide a list of medication names and attributes, which included dose, frequency, route, PRN (as needed) status and time stamps, if applicable. Their answers were classified into the following four categories:

1. *Correct and complete*, meaning the medication list was complete, and medication attributes such as dose and frequency were accurate.
2. *Correct but incomplete*, meaning the list of medication names was complete, but some or all of the medication attributes were missing.
3. *Omitted*, meaning certain medication names were missing from the list.
4. *Incorrect*, meaning either the list of medication names or medication attributes were erroneous.

Only answers in category 1, correct and complete, were considered *accurate*. Answers that fell into the other three categories were considered *inaccurate*.

For task 5, subjects were asked to provide drug class information for three medications taken by the given patient. Responses were graded as either *correct* or *incorrect*.

The average percentage of inaccuracy by different subjects for each case was calculated. For the same case, the percentages of inaccuracy under the two system circumstances were compared. Because each case had five tasks, the percentages of inaccuracy by different subjects on each task under the two system circumstances were calculated and compared. In other words, as shown in Table 4, comparisons were made between the following pairs under two system circumstances:

1. Average percentage of inaccuracy occurring during all tasks.
2. Average percentage of inaccuracy occurring when completing task 1—find out admission time.
3. Average percentage of inaccuracy occurring when completing task 2—find out inpatient medications.
4. Average percentage of inaccuracy occurring when completing task 3—find out medication history.
5. Average percentage of inaccuracy occurring when completing task 4—find out discharge medications.
6. Average percentage of inaccuracy occurring when completing task 5—find out medication classes.

Furthermore, comparisons were also made based on the inaccuracy types. In particular, comparisons were made for the following pairs under two system circumstances: (1) average percentage of omitted answers, (2) average percentage of incomplete answers, and (3) average percentage of incorrect answers.

2.7.2. Qualitative Analysis—The main qualitative analysis was to survey the overall impression from subjects regarding the clinical usefulness of the Timeline application.

Subjects gave scores to twelve Likert style survey questions, where one stood for strongly disagree and five stood for strongly agree. The average score and score distribution for each question were computed. Comments by subjects were collected and categorized into “ease of use,” “satisfaction,” and “clinical usefulness.”

2.7.3. Statistical Analysis—Wilcoxon matched-pair signed-rank test (Daniel, 1999) was used to determine the magnitude of differences between performance times (continuous variable) in using the two system environments. As an alternative to the paired t-test, this test was suitable for this study because the subject population could not be assumed to be normally distributed.

The difference between performance inaccuracy rates of using the two system environments was calculated using Fisher’s exact (Daniel, 1999), because the sample size was small (chi-square would be used if the sample size was greater than 10) and the variable used in this study, proportion of inaccurate responses, was categorical in nature.

Inter-rater reliability of the survey was calculated using Kendall’s coefficient of concordance (known as Kendall’s W) (Kendall and Smith, 1939) to measure the agreement among six subjects (raters) who were ranking a given set of twelve questions in a Likert style (ordinal) questionnaire. The reason why Kendall’s coefficient of concordance W was selected over Kappa was because Cohen’s Kappa was usually applied to measure agreement between two raters on nominal data, but there were more than two raters in the survey; and Fleiss’ Kappa was for assessing the reliability of agreement between a fixed number of raters when assigning categorical ratings to a number of items, but data in this survey study were ordinal. On the other hand, although good for ordinal data, Spearman Rho is often used as a non-parametric measure of correlation to describe the relationship between only two variables without making any other assumptions about the particular nature of the relationship between the variables. Therefore Spearman Rho was not a suitable choice for this analysis.

3. Results

Six male Internal Medicine residents volunteered to participate in the study. They were between PGY-1 (first year) and PGY-3 (third year) in their residency program at the Department of Medicine at NYPH. All of them had computer experience for more than ten years, and have been exposed to EHR systems such as WebCIS for at least three years.

3.1. Quantitative Analysis

3.1.1. Performance Efficiency—The Box Whisker plot (Figure 3) displays the comparison of times spent per subject on all tasks in two system environments—with and without the Timeline intervention. The x-axis shows the five tasks that all subjects performed on all records, and the y-axis shows the average time spent per subject per task measured in minutes. For each box, the upper and lower borders show the first quartile (25%) and third quartile (75%), and the middle line shows the median time spent on finish a task. The higher and lower ends of the vertical line pertaining to the box display the maximum and minimum time spent for that task.

Statistically, the time difference between using WebCIS only and WebCIS plus Timeline proved to be significant on the following tasks by using non-parametric paired test (Wilcoxon signed-rank test with $\alpha=0.05$) admission time (p-value=0.028), inpatient medications (p-value= 0.028), pre-admission medications (p-value=0.028), and medication classes (p-value= 0.046).

The following Table 5 shows the time difference across all records among all subjects between using WebCIS only and using WebCIS plus Timeline. For example, time spent for the admission time task was noted as 1.11 minutes for the WebCIS group and 0.08 minutes for WebCIS plus Timeline, yielding a difference of 1.03 minutes. Overall, the average time spent on all tasks was 4.91 minutes using WebCIS, and 0.92 minutes using WebCIS plus Timeline, with nearly a four-minute difference per record. Among all the recorded times, the longest was from subject 4 when trying to finish the inpatient medication task on Scenario 5. He spent close to eight minutes to complete, whereas most of other subjects spent no more than three minutes on the same task, which in part led to the highest standard deviation in this table. The least time-consuming task in this study was using WebCIS plus Timeline to find the admission time. Subjects all spent a similar amount of time to complete this task, which was reflected in the low standard deviation as well.

Among all the tasks finished using only WebCIS, subjects spent from the longest to the shortest time on inpatient medications, admission time, medication classes, and pre-admission medications, followed by discharge medications. For the same tasks that subjects accomplished using WebCIS plus Timeline, the proportion of time spent on each task was different. Subjects spent from the longest to the shortest time on finding medication classes, followed by preadmission medications, inpatient medications, discharge medications and admission time. The two sets of time spent on the same list of tasks yielded a new order for the time difference between using the two system environments. Since there were no negative numbers in the difference column, one could conclude that for all tasks, using WebCIS plus Timeline saved some time over using only WebCIS. The task that showed the biggest difference implied the most time-saving effect of the Timeline intervention. As shown in the table, the task with the greatest time-saving effect was inpatient medications, followed by admission time, preadmission medications, medication classes, and discharge medications.

3.1.2. Performance Inaccuracy—The gold standard process described generated a medication list based on consensus (Kappa = 1) with information on medication names, dosages, routes, frequencies, PRN status, start/end time if available, and data sources. For each scenario, the patient's last inpatient admission date was identified and documented. Table 6 shows the number of medications identified and documented for each scenario upon inpatient admission, pre-admission, and discharge. Generating the reference list required several iterations to reconcile the difference between two clinicians. In the earlier states of this process, most disagreements were related to start/end time information of certain medications due to the lack of sufficient temporal information in the system, but the experts eventually reached agreement on their best estimates based on available data. Medication classes were retrieved from the American Hospital Formulary Services Drug Information (AHFS-DI) database and confirmed by experts.

Figure 4 shows the difference between average percentages of inaccuracy in completing the five tasks in two system settings, where the x-axis represents the tasks performed by subjects, and the y-axis represents the inaccuracy rate (the average percentage of inaccurate answers per subject per task). On average, most tasks were finished with less than 60% inaccuracy percentage. The visual gaps were evident between those of inpatient medications, admission time and pre-admission medications when comparing the two system environments.

The inaccuracy rate was the number of inaccurate answers divided by the number of gold standard answers. For example, in Scenario 1, if a subject found two inpatient medications where the gold standard reference list was nine, his percentage of inaccuracy was $(9-2)/9=77.8\%$.

Among all the tasks finished in WebCIS, the highest inaccuracy rate appeared in finding inpatient medications (54.76%), which meant that on average, more than half of the answers provided by subjects on inpatient medications were inaccurate. Other high percentages of inaccuracies occurred in finding pre-admission medications and admission time. The most accurate results came from identifying medication classes and discharge medications. The latter was also the task that subjects spent the least amount of time on as shown in Table 5.

For all the tasks completed in WebCIS plus Timeline, the highest inaccuracy rate was found in determining pre-admission medications (16.67%), which meant that on average, one sixth of the answers provided by subjects on pre-admission medications were inaccurate. The next highest inaccuracy rate was found in finding discharge medications, followed by inpatient medications. No subjects made mistakes in identifying admission time and medication classes, each of which scored 0% inaccuracy rate.

The difference between inaccuracy rates in the two system settings shows that with the intervention of Timeline, the average inaccuracy percentage of subjects' answers can be improved over WebCIS alone to some degree on four tasks (except for medication classes). The implication is that the greater the difference, the more improvement in the inaccuracy rate one could expect. In other words, the task that showed the biggest difference would imply the most inaccuracy-reducing effect by the intervention of Timeline. For example, one can potentially reduce the percentage of inaccurate inpatient medications by more than 50% with the help of Timeline.

As discussed in Section 2.7.1.2, for each inaccurate answer on inpatient, pre-admission, and discharge medications, there were three possible states—omitted, incomplete, or incorrect (erroneous) medication. For each answer on admission time and medication classes, the result was either correct or incorrect. Although they each contributed to the total number of inaccuracies, different types of inaccuracy would have dissimilar clinical impact and consequences if they occurred. Therefore, further analysis was conducted on each possibility.

Table 7 summarizes and compares the average percentages of inaccuracy including omitted, incomplete and incorrect answers on admission time, inpatient, pre-admission, and discharge medications, and medication classes between using WebCIS only and WebCIS plus

Timeline. The differences of error rates on admission time and omission rates on inpatient medications were statistically significant according to Fisher's exact test, and the p-values were 0.045 and 0.003 respectively. Among all inaccuracy rates, those greater than 15% all appeared under the "omitted" column. The percentages of incorrect medications were the lowest, all of which were less than 6%. Percentages of incomplete medications fell in between, ranging from 0% to 13%. Because this study did not apply a weighting mechanism to further calculate the importance score of each type, each type weighed the same towards the final inaccuracy rate. Therefore, the distribution of each inaccuracy type implies how much that type contributed to a high inaccuracy rate; for example, for inpatient medications using WebCIS, high number of omitted and incomplete medications caused a high overall inaccuracy rate. Put differently, by reducing the high omission and incompleteness rates on inpatient medications, the total inaccuracy rate can be improved to a large extent. The results indicated that the bigger the difference between two system environments, the more potential there was for Timeline to help reduce the inaccuracy rate.

3.2. Qualitative Analysis

3.2.1. Survey—In this section, survey results include: (1) average survey scores calculated based on original individual scores, (2) score distribution, and (3) the degree of agreement from all subjects on all questions. Figure 5 shows average scores from all subjects on all survey questions, where one stands for strongly disagree and five stands for strongly agree. Figure 6 shows the degree of agreement among subjects.

The full survey questionnaire can be found in Appendix.

The first question in the survey assessed the extent to which the prototype Timeline solution provided important medication information to users in general. Most subjects rated it favorably, with an average rating of 4.33 out of 5.

Questions 2–7 assessed the clinical usefulness of Timeline by asking if users found its functions and features helpful in practice, such as its temporal and visual representation, medication classification, and data sources. Most subjects gave high scores on those questions, except for subject 2, who was critical and rated two questions 2 out of 5.

Questions 8 and 9 evaluated the users' perception on whether the Timeline intervention could help improve efficiency and accuracy in their work. Most subjects except for subject 2 responded positively and the average scores were 4.17 and 4.83. It appeared that most subjects felt more confident in improving accuracy than saving time with the help from Timeline.

Question 10 assessed the ease of use of Timeline, and the average score was among the highest from the survey. Basically all subjects found the application very easy to learn and to use.

Questions 11 and 12 assessed the satisfaction level from the users. Subjects were asked if they found Timeline application useful enough that they would want to use and recommend

it to peers and colleagues at work. Answers on questions 11 and 12 gave the highest and second highest scores of the survey respectively.

In summary, as shown in Figure 5, average scores on all questions were greater than 4.1. The highest average score (4.83) was given to three questions: ease of use of Timeline, confidence in Timeline to improve accuracy, and satisfaction with overall Timeline usefulness. The lowest average score (4.17) was given to the following two: using Timeline to improve efficiency and to identify medication duration in all clinical settings.

Besides average score on each survey question, average scores for each subject were calculated. The highest average score (4.92) was given by subject 1 and 6, and the lowest average score (3.75) was given by subject 2.

The quantitative calculation of degree of agreement, or similarity in subjects' answers was shown in Figure 6. The inter-rater reliability resulted in 0.305 using Kendall's coefficient of concordance, with subject 2 having the highest standard deviation and lowest mean rank, and subject 1 having the lowest standard deviation and the highest mean rank. This results suggested moderate agreement among subjects.

Analyses on score distribution from all subjects on all questions showed that there were three incidents where subjects scored 2 out of 5 on certain questions (4.1%), and two incidents where subjects scored 3 out of 5 (2.8%). Subjects gave 5 out of 5 to 47 questions (65%) and 4 out of 5 to 20 questions (27.8%). This results indicated that the majority survey questions (more than 92%) received 4 or 5 out of 5 from subjects.

In the free-text comment field, some subjects wrote down their impression of the proposed Timeline application and additional comments or concerns that were not covered by the survey. Subject 1 said "I think this application is very useful in terms of looking at a list of medications that may have been compiled from several clinical sources." Subject 3 stated "(Timeline application) makes life easier!" Subject 6 wrote "It fills an important need for a centralized source of information."

Subject 5 commented on the user interface during the interview: "I like it a lot. I like how visual it is. It is quite user friendly. I made comments about things that could be nice to have, but I mean it's not difficult to use this at all." There were also many suggestions and recommendations on Timeline functions and user interface. In particular, Subjects 2 and 6 recommended to "add copy-paste function to the application". Subject 6 would like to see an "expanded view" in addition to the current representation so that more regimen details could be displayed. Subject 2 preferred to have a cleaner and simpler view of the most important information, and suggested to consolidate same medications into one line regardless of their data sources and vary the thickness to represent changes on dose. Subject 5 preferred the opposite and said "I actually like it that they were sort of set apart [pointed at the medication history and outpatient medications]. I was fooled that this [pointed at outpatient medication section] was all the way down here, but it was nice!"

When asked if there were any challenges in using Timeline, Subject 1 answered "I don't have any complaints. I think if I knew how to use it better that I would be able to [perform

better]. For instance, I'm not entirely sure the difference between outpatient drugs and the med hx. So if I were looking for what drugs the patients were on before they came to the hospital, I wouldn't be sure. Does that mean all of them? I think if I just get used to it [Timeline] and actually learn what each of them [Timeline labels] mean, it would be very clear. Otherwise the interface was great." Subject 5 said "Overall it is such an improvement over what I was trying to do before. It's good to have such a reliable source."

4. Discussion

An important point of this study was that Timeline was not brought into the experiment to replace any existing EHRs; rather, it was designed to work as a complementary summarization tool to help clinicians gather important medication information. Timeline did not cover information beyond medications, and did not repeat ordering functions most modern Computerized Physician Order Entry (CPOE) systems provide. Clinicians would still use the systems they are familiar with, such as WebCIS, to write notes, but when reviewing patients' medication information, they could either first take a quick look at the Timeline application and then refer back to the original notes, or summarize medications in their own way and use the Timeline application for confirmation. As one subject (S2) pointed out when asked if Timeline could help him generate a more accurate medication list, "Not alone" he said, because there was an enormous amount of information scattered in various notes, even though Timeline included excerpts from notes, being able to go back to the original context was still critical for most subjects to create an accurate medication list. Therefore, the Timeline application is supposed to work as a plug-in application in addition to WebCIS or similar systems, instead of a full-fledged standalone medication reference tool.

Given the scale of the study, statistical analysis could be best viewed as exploratory instead of the most prevailing measure to evaluate the study results. Nonetheless, the results of the statistical tests applied in this study could help provide additional quantitative insight into the overall evaluation that cannot be elicited from qualitative evaluation performed.

With the intervention of Timeline, there was a significant time reduction for most tasks carried out in this study except for task 4, "finding discharge medications." This was in large part due to the fact that without Timeline's help, most subjects did not spend much time completing the task in the first place, even though they did spend less time on this task when receiving assistance from Timeline. This result was not surprising. As discovered in the cognitive study on understanding clinicians' mental process of medication summarization, finding discharge medications was one of the most streamlined and straightforward processes for most clinicians. On the other hand, more complex processes, such as finding pre-admission medications, showed an obvious difference in terms of completion time with and without the intervention of Timeline. However, regardless of overall positive results from the statistical analysis, subjects' perception on whether the Timeline intervention would help them improve their efficiency was more conservative. One concern was about the time to learn and to get used to a new user interface. During the orientation, Subject 3 said "if you didn't show me how to use, I'll have to spend more time to figure it out by myself". A proper and common solution to this is to introduce adequate orientation or

training to potential users, which is generally effective and necessary for most software applications in today's information technology field. Other concerns were related to features of Timeline. Subject 6 thought having a "cut and paste" function would save time, and Subject 2 considered a simpler graphical user interface would improve efficiency because "it [reviewing medications] is difficult especially when a patient is complicated and had a prolonged stay". These were valuable suggestions for future implementation.

Subjects gave the highest agreement score in the survey when asked if they thought Timeline could reduce inaccuracy in their practice. It was encouraging to see great confidence from subjects, and it indicates the inaccuracy rate can be further reduced when (1) clinical users get more familiar with Timeline, (2) the algorithms used to provide Timeline content are improved, and (3) original data source become more accurate. There was a case (Scenario 5) in this study in which mistakes from the original data source caused errors in Timeline, which later confused subjects. In the patient record of Scenario 5, after the patient died, there were still two more medication orders placed by care providers possibly by mistake in the pharmacy order system. It was narrated in the notes that the patient was expired, and there was no discharge summary for the patient's last hospitalization. The medication Timeline application used medication data from both narrative text notes and structured database records; for the narrative notes, it used the previously published medication parser (Gold et al., 2008) to help retrieve discharge medications from free-text discharge summaries. But in this case, because there was no discharge summary, no discharge medications were retrieved from notes. For structured medication data, Timeline queried the pharmacy database and then displayed the two mistakenly ordered medications in the area for discharge or outpatient medications. Timeline represented what was retrieved from databases, but also inherited the mistakes from them. Timeline algorithms were not intelligent enough to correct the mistakes in the database with knowledge from the notes. On the other hand, when subjects were only using the notes system, they read "expired" in the notes, and made the right conclusion immediately that there was no discharge medication from the last hospital stay. Although the root cause of this issue was due to the current quality of medical notes and orders, there is room for Timeline to improve and be more useful. An intuitive solution to this problem is to create an alert mechanism, which can be triggered when medications from various data sources were different. Another possible solution is to train medication parser to be able to recognize context key words such as "expired" in notes, and then implement decision support logic to detect potential mistakes, for example, "if a patient is expired → find expiration day → there should be no medications after that date → if there were medications after the expiration day, send alarms".

Another interesting finding was that even without the help from Timeline, subjects did not spend too much time on finding medication classes and their answers bore minimal errors. However, they did spend more time and made more errors on finding inpatient and preadmission medications, but the results were improved by the intervention of Timeline. This was in concert with findings from the cognitive study, where processes on finding inpatient and preadmission medications were the most complex and time-consuming process, implying that the results of those processes might be the most error-prone and users may need more assistance.

One potential explanation for this finding on medication classes is based on the assumption that human beings need more assistance for tasks that rely extensively on working memory, but need less help if the knowledge required to perform certain cognitive activities resides in long-term memory and has been reinforced since the acquisition. Specifically, subjects had learned what medications belonged to which medication classes through their medical training many years ago, and reinforced that knowledge in daily practice. Therefore, when using WebCIS, they answered questions from “things learned by heart,” which did not take them too much effort to accomplish. When given Timeline as an option, most subjects used it as a reference tool to confirm what they already knew and had in their minds. This finding was in line with observations from the cognitive study, where one of the subjects pointed out explicitly that she would think most residents at CUMC should be able to identify medication classes well, even though it would be a nice feature to have.

On the other hand, there were more than 140 inpatient and pre-admission medications in this study that each subject had to process within half an hour. Most subjects relied on their working memories to gather information from various data sources, and to remember and process information about these medications. According to cognitive theories published by (Patel and Kaufman, 2006), human working memory is normally limited to 5 to 10 “chunks” of information. The information maintained in working memory includes knowledge activated from long-term memory, and stimuli from the environment, such as what was displayed on the screen. An excess of information that competes for scarce cognitive resources can burden working memory. Multitasking, scattered data sources, and unfriendly user interface can be contributing factors for cognitive overload. In this study, most subjects could name three medication classes with no problem by activating and transferring knowledge from their long-term memory to working memory, but when they were asked to search and process dozens of medications from many resources for inpatient and pre-admission medications, the cognitive load on working memory increased significantly. This is similar to the example that many people do not have trouble maintaining information on three restaurants, but would need help on maintaining names, locations, or phone numbers of dozens of restaurants while engaging in an online search or a lookup in the yellow pages. Most people would need either to summarize a list of their search results or to highlight what they find on different pages. Timeline was designed according to these cognitive theories, hoping to lighten some if not all of the burdens on clinicians’ working memories by imitating their thought process and preparing a summary list for them. The results of the study show both the promising effect on execution time and error rates, as well as appreciation from clinicians on the design and functions of the Timeline application. As Subject 5 commented “I’m glad to see that the data were there. Now I feel like I’m done, I don’t even feel like looking at other places, because this [pointed at Timeline] looks like the kind of list that I would expect. So it sort of fits intuitively.”

Overall, the survey results were gratifying, but the inter-rater reliability showed some disagreement among subjects. Some subjects had differing views on issues from user interface to software function, as described in Section 3.2.1. The divergence in opinions was possibly because of different personal preferences and experiences. For example, we observed that the subject (Subject 2) who had the most experience in the residency program (more than 3 years) appeared to see more problems and challenges than others, and the

subject (Subject 5) who worked as a computer engineer before enrolling into medical school appeared visibly more optimistic and positive than the rest of the subjects.

Clinically, some subjects would like to be able to distinguish important medications from those with less clinical significance, for example, to be able to distinguish the clinical impact of morphine from that of multivitamin or docusate. On a broader level, this is feasible. However, this could be a tricky and challenging task at a finer scale; for instance, it may be difficult to tell whether a beta blocker was more clinically important than a pain reliever, depending on if they were for a hypertensive patient with menstrual discomfort or terminal cancer. A possible solution is to include a simple scoring system to rank the clinical importance of frequently used medications in basic groups, but also to provide an “advanced” function to allow clinicians to redefine the level of clinical importance on an ad hoc basis. Then, with the current filter function in Timeline, clinicians can view medications at the appropriate level of clinical importance.

The results of this research were subject to several limitations. First, the evaluation study was conducted based on simulations of clinical environments using a prototype application. More advanced evaluations can be done on a larger scale after integrating the proposed medication summarization tool into real-time clinical information systems. Second, a series of medication summarization processes was implemented in a general clinical sense. In other words, this work was based on the examination of general needs of a small size of clinical users. In practice, different clinical user groups may have different information needs, and a user with certain needs in one situation might have different needs in another situation. Subjects recruited for the studies in this project do not represent particular practicing groups with specific information needs, and how to incorporate users’ specific needs or/and dynamic changes in situations was not within the scope of this research. Overall, the proposed Timeline application is a general solution for any EHR that exploits available data sources to supplement the provider’s summarization of an accurate medication list. The results of this study can be generalized and applied to hospital systems that intend to extract, merge and represent medication information from both narrative clinical notes and structured electronic health records. At different interfaces of care, as long as the data sources are in those two forms, this approach is applicable. However, limitations exist when medication information has been recorded and stored in scanned images, pictures, pill samples, and un-transcribed audio tapes.

5. Conclusion

The overall result of this evaluation study was satisfying. Results from time analyses showed clear differences between times spent with and without the intervention of Timeline on the following tasks: admission time, inpatient medications, pre-admission medications, and medication classes. Results from inaccuracy rate analyses showed significant improvement with the intervention of Timeline on admission time and inpatient medications. Survey results showed positive feedback on Timeline regarding its clinical usefulness, ease of use and user satisfaction. The results of this study demonstrated great potential and feasibility of implementing such a medication summarization tool into clinicians’ daily practice in addition to existing EHRs to improve the efficiency and

accuracy of their performance. Future work includes improvement in the graphical user interface and data integration.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

This study was supported by National Library of Medicine RO1 LM006910 and T15 LM007079. The authors thank Dr. David Kaufman and all study participants for their valuable comments and insights.

References

1. Barnsteiner JH. Medication reconciliation: transfer of medication information across settings-keeping it free from error. *American Journal of Nursing*. 2005; 105(3 Suppl):31–36. quiz 48–51. [PubMed: 15802996]
2. Daniel, WW. *Biostatistics: A Foundation for Analysis in the Health Sciences*. Wiley; 1999.
3. Gold, S., et al. Extracting Structured Medication Event Information from Discharge Summaries; AMIA ... Annual Symposium Proceedings/AMIA Symposium; 2008.
4. Hripcsak, G., et al. WebCIS: large scale deployment of a Web-based clinical information system; AMIA ... Annual Symposium Proceedings/AMIA Symposium; 1999. p. 804-808.
5. IHI. Reconciling Medications to Avoid Medical Errors. 2009. from <http://www.ihl.org/IHI/Topics/PatientSafety/MedicationSystems/ImprovementStories/ReconcilingMedicationsToAvoidMedicalErrors.htm>.
6. Kendall MG, Smith BB. The Problem of m Rankings. *The Annals of Mathematical Statistics*. 1939; 10(3):275–287.
7. Morae. Usability testing and market research software. 1995–2009. from <http://www.techsmith.com/morae.html>.
8. Patel, VL.; Kaufman, DR. Shortliffe, EHL.; Cimino, JJ. *Cognitive Science and Biomedical Informatics Biomedical Informatics, Computer Applications in Health Care and Biomedicine*. New York: Springer; 2006. p. 133-185.
9. Timeline. SIMILE Web Widget for Visualizing Temporal Data. 2006–2009. from <http://www.simile-widgets.org/timeline/>.
10. Zhu X, et al. Using Timeline Displays to Improve Medication Reconciliation. *International Conference on eHealth, Telemedicine, and Social Medicine. eTELEMED '09*. 2009:1–6.

Fig. 1.
Medication Summarization Timeline Application.

Fig. 2.
Gold Standard Generation.

Fig. 3.
Time Spent on All Tasks with and without Timeline Intervention.

Fig. 4. Average Percentage of Inaccurate Answers per Subject per Task in Two System Environments.

Fig. 5.
Average Scores from All Subjects on All Survey Questions (1– strongly disagree 5– strongly agree).

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles		
						25th	50th (Median)	75th
observer1	12	4.9167	.28868	4.00	5.00	5.0000	5.0000	5.0000
observer2	12	3.7500	1.21543	2.00	5.00	2.2500	4.0000	5.0000
observer3	12	4.3333	.49237	4.00	5.00	4.0000	4.0000	5.0000
observer4	12	4.6667	.49237	4.00	5.00	4.0000	5.0000	5.0000
observer5	12	4.6667	.65134	3.00	5.00	4.2500	5.0000	5.0000
observer6	12	4.9167	.28868	4.00	5.00	5.0000	5.0000	5.0000

Kendall's W Test**Ranks**

	Mean Rank
observer1	4.33
observer2	2.25
observer3	2.71
observer4	3.63
observer5	3.79
observer6	4.29

Test Statistics

N	12
Kendall's W ^a	.305
Chi-Square	18.310
df	5
Asymp. Sig.	.003

a. Kendall's
Coefficient of
Concordance

Fig. 6.
Degree of Agreement among Subjects.

Table 1

Clinical Tasks for Clinicians during Patient Chart Review.

Tasks
1. When was the patient's most recent hospital stay or inpatient admission?
2. What medications did the patient take during his/her most recent hospital stay?
3. What medications had the patient been taking before the latest admission to treat his/her conditions?
4. What were the discharge medications?
5. Name three drug therapeutic classes the patient had been put on

Table 2

Patient Records Assignment.

Records	Subject A	Subject B	Subject C	Subject D	Subject E	Subject F
I: Without Med Timeline	1 2 3	2 3 4	3 4 5	4 5 6	5 6 1	6 1 2
II: With Med Timeline	4 5 6	1 5 6	1 2 6	1 2 3	2 3 4	3 4 5

Table 3

Sample Survey Questions.

Please rate the following questions on an 1–5 scale, where 1 is "strongly disagree" and 5 is "strongly agree":
This application presents the important medication information I want to know
I can see when a medication was started
I can tell the duration of time what a medication has been in use
I can conclude when the medication was discontinued
I can tell which drug class a medication belongs to
I can view medication based on different data sources
This type of application will be useful in practice
I find this application useful because it could help me generate a more accurate medication list

Table 4

Data Analysis of Inaccuracy Percentage on All Tasks.

Tasks	Omission Rate		Incompletion Rate		Error Rate		Total Inaccuracy Rate	
Admission time	Web CIS	WebCIS+Timeline	Web CIS	WebCIS+Timeline	Web CIS	WebCIS+Timeline	Web CIS	WebCIS+Timeline
Inpatient Medications								
Pre-admission Medications								
Discharge Medications								
Medication Classes								

Table 5

Average Time Spent on Performing All Tasks in Two System Environments (Time measured in minutes).

Task	Time Spent (SD)		Difference	p-value
	Without Med Timeline	With Med Timeline		
Admission Time	1.11 (0.65)	0.08 (0.05)	1.03	0.028*
Inpatient Medications	1.96 (1.98)	0.16 (0.09)	1.8	0.028*
Pre-admission Medications	0.69 (0.28)	0.23 (0.18)	0.46	0.028*
Discharge Medications	0.41 (0.37)	0.12 (0.1)	0.29	0.075
Medication Classes	0.74 (0.11)	0.33 (0.26)	0.41	0.046*
Total	4.91	0.92	3.99	

* (p<0.05) Wilcoxon signed-rank test.

Table 6

Medications in Reference List.

Tasks	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Total
Inpatient Medications	9	7	6	23	35	18	98
Pre-adm. Medications	7	0	4	11	14	8	44
Discharge Medications	4	3	6	0	0	11	24
Total	20	10	16	34	49	37	166

Table 7

Average Percentage of Inaccurate (Including Omitted, Incomplete and Incorrect) Answers on All Tasks by All Subjects in Two System Environments.

Rate	Omitted		Incomplete		Incorrect		Total Inaccurate	
Tasks	WebCIS	WebCIS + Timeline	WebCIS	WebCIS + Timeline	WebCIS	WebCIS + Timeline	WebCIS	WebCIS + Timeline
Admission time	N/A	N/A	N/A	N/A	27.8% *	0% *	27.80%	0%
Inpatient Medications	42.18% *	0.34% *	8.50%	2.04%	4.08%	0.34%	54.76%	2.72%
Pre-admission Medications	17.42%	15.15%	12.88%	1.52%	0.76%	0%	31.06%	16.67%
Discharge Medications	12.50%	5.56%	0%	0%	0%	5.56%	12.50%	11.12%
Medication Classes	N/A	N/A	N/A	N/A	0%	0%	0%	0%

* (p<0.05) Fisher's exact test.