
This project designed and implemented visualization interface of EHR Systems based on the requirements of the doctors. The visualizations combined both clinic-reported data and patient self-reported data to provide a better representation of patient health related information for the doctor to make decisions. The visualizations highlighted the trend of values, the outliers and make it possible to compare across time and measures. The user study of ten participants suggests that the visualization interface helped them find the information in an efficient way.

Headings:

Visual Analysis
EHR
Flowsheet
Health-related Data
Patient Self-reported Data
A VISUAL ANALYSIS OF EHR FLOWSHEET TO ASSIST CLINICIANS’ INTERPRETATION OF HEALTH-RELATED DATA

by
Lu Xu

A Master’s paper submitted to the faculty of the School of Information and Library Science of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in Information Science.

Chapel Hill, North Carolina
April 2019

Approved by

_______________________________________
David Gotz
Table of Contents

**INTRODUCTION** .......................................................................................................................... 3

**LITERATURE REVIEW** .................................................................................................................. 6

1.1 Doctor Interaction with EHR Systems .......................................................................................... 6

1.2 Data Visualization ......................................................................................................................... 7

1.3 Data Analysis .............................................................................................................................. 7

1.4 Brief Summary ........................................................................................................................... 8

**METHODS** .................................................................................................................................. 9

1. Data Collection .............................................................................................................................. 9

2. Research Design ........................................................................................................................... 11

2.1 Data Analysis and Visualization .................................................................................................. 11

2.2 Evaluation .................................................................................................................................. 12

2.2.1 Survey ................................................................................................................................... 12

2.2.2 Semi-structured Interview ...................................................................................................... 13

3. Limitations ..................................................................................................................................... 14

**SYSTEM DESIGN** ......................................................................................................................... 15

1. User Requirement .......................................................................................................................... 15

2. System Functions .......................................................................................................................... 15

2.1 Measure Types ............................................................................................................................ 16

2.2 Measure Display ........................................................................................................................ 16

2.3 Self-reported Measures .............................................................................................................. 17

**USER STUDY AND ANALYSIS** .................................................................................................. 19

1. Participants ................................................................................................................................... 19

2. Analysis of Survey Results ........................................................................................................... 20

3. Limitations and Future Work ....................................................................................................... 21

**CONCLUSION** .............................................................................................................................. 22
INTRODUCTION

Diabetes Mellitus is a chronic disease. Management for Diabetes is concentrated in keeping blood glucose level close to normal. For this purpose, monitoring and control blood glucose, weight, pulse and other vital signs to a health level are important.

The current Electronic Health Records (EHR) systems that UNC Hospital are using to record patients’ vital signs and other measures includes two dimensional tables of variables and value pairs. The patient’s measure values of each visit to the clinic/hospitals are recorded in the EHR systems, and the patient self-generated health data are not collected in the systems. The medicine records and measures are in different tables and are not well connected with each other, so it is almost impossible for clinicians to check the medicine records and the corresponding measures after each medicine record during the patient visit time (around 15 min).

In this project, I am interested in data visual analysis of EHR Systems for diabetes patients. Two types of EHR are visualized and analyzed: the clinical records and patients self-generated data. Patient self-reported data is generally more frequent than clinical-reported data. Patients can track their blood pressure, temperature, pulse, weight and other measures of their body’s basic functions once or multiple times a day by a simple wearable or fixed equipment. Clinical data, however, can only be measured/recorded
when patients visit clinics. If the equipment can upload data automatically and/or user can also type in the values, the information provided by both clinical visit data and patient-self reported data can be much more informative for doctors to understand the patient’s status.

The client group of the proposed visualization interface is doctors with M.D. degree, who work in emergency as well as non-emergency environment in the hospitals or clinics, and are using current EHR systems (such as Epic) to record patients’ health related information.

The research question for this project is how to visualize health related measures of diabetes patients, combined both patient self-reported and clinical reported information, to assist with clinicians’ understanding of these data and help them to make decisions.

The following are the questions about patient self-report data and clinic-reported data need to be figured out before the design of visualizations:

1. What measures are important for analysis and are visualized in the chart for the clinicians to view?

2. Both patient self-reported data and clinical data have different frequency, different categories, different ranges and different measures. Can they be displayed in the same chart?

3. If they can be displayed in the same chart, how is the time axis divided? (use the time point of the patient visit clinic?)
4. Are there some good examples to visualize them?

5. How to make unusual values obvious for clinicians to find out?

After meetings with Dr. Arlene Chung and Dr. David Gotz, we selected the measures that most important for the diabetes patients. Dr. Chung is a UNC Lineberger Comprehensive Cancer Center member and an Assistant Professor of Medicine and Pediatrics and Associate Director of Health and Clinical Informatics for the UNC School of Medicine. The variables we decided to visualize are Blood Glucose (including Glucose POC (fingerstick) and Glucose Bld (blood draw)), Vital Signs (Weight, Blood Pressure, Pulse, Pulse Oxygen), and Medicine. After the design and implement of the visualizations, I recruited ten persons with M.D. degree or with B.S / B.A degrees to participate in user experiment on clinicians and other participants to evaluate the visualization interface of information systems that I designed and implemented, in order to find out whether the visualizations are able to help doctors understand the patient’s health-related information.
LITERATURE REVIEW

There are several fields which have relevant studies, such as Electronic Health Records (EHR), how to visualize health-related data, doctor interaction with EHR, and data visualization. Following are the theories and opinions which are helpful and necessary to look at.

1.1 Doctor Interaction with EHR Systems

Montague (2013) mentions that EHR improves patient’s satisfaction. However, the results show that technology-centered physicians spent the most time inputting and extracting information, which may limit their time to interact with the technology at high levels or recall information about the patient. For this situation, find an efficient way to help improve the interactions between the doctors and EHR Systems are necessary.

Benjamin et al. (2018) mentions that EHRs are often perceived by users as difficult to use. There are issues with difficult-to-read interfaces, confusing displays and iconography that lacks consistency and intuitive meaning. These usability issues can often lead to heavy cognitive load.

The difficulties that EHR currently have are hard-to-read interfaces, confusing displays that lacks of intuitive meaning, and requiring more time to extract the data required.
Correspondingly, the information needs of the clients are simple, efficient access of the patient health related information to help them to have a better understanding of the patients and distinguish some outliers and abnormal patterns of the patients’ vital signs.

1.2 Data Visualization

Data visualization is a very broad concept, and according to the definition by Azzam et al. (2013), it is a process which is based on qualitative and quantitative data and results in an image that represents the raw data, and the image is readable by viewers and supports exploration, examination and communication of the data, while the theory of this definition is originally from Kosara (2007).

While the technique of data visualization gets much developed and advanced, there are still concerns and challenges of this field. Kovalerchuk (2017) studies the lossless visualization of n-D data and considers it better than 2-D in perceptual and cognitive abilities for discovering visual pattern, but human are limited in their ability to discover patterns in n-D data using a naked eye, which therefore motivates the development of 2-D visual representations of n-D data.

1.3 Evaluation of Visualization

The evaluation of information visualization is necessary because the visualization interface is designed and developed to meet the users’ requirements.
Carpendale (2008) mentioned that usability questions, perceptual and comprehensibility questions are important in assessing the appropriateness of a representational encoding and the readability of visuals.

Studies on data visualization and visual analytics in health informatics applications commonly use task-based measurements, user feedback, and surveys with both qualitative responses and scoring scales in a lab setting (Gotz et al, 2019). The PIOS framework (Participants, Interventions, Comparators, Outcomes, and Study Design) proposed by Gotz et al. also provide a framework for this project.

The evaluation methods that I applied are task-based questions, surveys about how user feel about the visualization interface and semi-structured interview to effectiveness, efficiency and satisfaction.

1.4 Brief Summary

Current research has already discussed about the difficulties of current EHR Systems, advantages of 2-D data visualization over n-D data visualization for human, and evaluation of visualization approaches for health informatics, which is very informative for the basic context of this project. But there are very few papers for visualizations of EHR systems, especially how to visualize both clinical data and patient self-reported data.
METHODS

This project includes both quantitative and qualitative methods for design and constructing the visualization interface and user study for visualization interface evaluation.

1. User Interview

To collect the user requirements for this project, Dr. Chung who is experienced with diabetes patients are interviewed. The questions answered include what data is needed, where the data is from, what kinds of insights are needed by a doctor.

2. Data

Data used for the visualization is mostly artificial dataset. The real patient clinical data are not considered because it is very sensitive. The patient-self reported data is not currently available in EHR systems, and it is the future features adding to the current EHR systems to help inform doctors about what happens between infrequent doctor visits, so that they are not available for this study. The examples online and Dr. Chung who has a good knowledge of the range of the normal values helped me to improve the artificial data quality and made it more realistic for visualization development and user evaluation.
The periodic clinical data with doctor’s office was also captured in today’s modern flowsheets, and user-generated data captured with high frequency in between visits.

The normal range of measures that included in the visualizations are as follows:

- **Weight**

  Weight has no upper or lower limits because the weight does not change dramatically. Thus, there are no boundaries and outliers for weight in the visualizations.

- **Pulse**

  Normal pulse rates at rest, in beats per minute (BPM):

<table>
<thead>
<tr>
<th>newborn (0–3 months old)</th>
<th>infants (3 – 6 months)</th>
<th>infants (6 – 12 months)</th>
<th>children (1 – 10 years)</th>
<th>children over 10 years &amp; adults, including seniors</th>
<th>well-trained adult athletes</th>
</tr>
</thead>
</table>

  This project selected 59 and 99 (children over 10 years & adults, including seniors) as lower and upper limit to find out outliers.

- **Blood Pressure**

<table>
<thead>
<tr>
<th>Category</th>
<th>systolic, mmHg</th>
<th>diastolic, mmHg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>90–119</td>
<td>60–79</td>
</tr>
<tr>
<td></td>
<td>90–129</td>
<td>60–84</td>
</tr>
</tbody>
</table>
• **Pulse Oxygen**

Oxygen saturation (SO2): Normal range is 75-100 mmHg. Below 60 mmHg indicates needs for supplemental oxygen. Values under 90 are considered low.

• **A1C**

Normal Range are between 5.7 - 6.4 percent.

• **Glucose**

A fasting blood glucose is in between 100-125 mg/dL. Any value between 140 mg/dL and 199 mg/dL during a two-hour 75g oral glucose tolerance test. Glucose is the same in terms of units/scale for both POC and BLD. The only difference would be if showing the hgA1C which is a percentage with an associated average rage of glucose with it.

• **INR**

This measure is weekly to a month if done at home; done after change in medication usually. INR below 1.1 is normal. Often INR intentionally kept higher (e.g., 2.0-3.0 for people with afib). High INR means blood will clot more slowly.

• **Medication**

The visualization interface added medication of each doctor visit.

3. **Research Design**

3.1 **Data Analysis and Visualization**

The main contribution of this project will be a visualization interface to show both the values of clinic-reported health-related data and patient self-reported data, so that the data analysis will be mainly about extracting and filtering data by different measures, finding
out the abnormal values and trend, then visualize them in the interface. The detailed description of system design is in the next section.

Other than cleaning the data and finding out the attributes needed for visualization, the design of visualization/interface is also one of the essential parts, which is followed by actually working on the code.

What’s more, after building the visualization, some analysis will also be included as illustrating findings and trends of this visualization, mainly for the purpose of identifying abnormal values and outliers.

3.2 Evaluation

As for constructing a data visualization project, it is necessary to have a final part as evaluation in order to test the efficiency of using this visualization. For this project I conducted questionnaire and semi-structured interview to evaluate the design of the interface and collect people’s opinions and suggestions while using the visualization interface. The evaluation of visualization will be in the User Study and Analysis section.

3.2.1 Participants

Subjects of the evaluation are the target users of the project – clinicians in UNC Hospital. The size of user study group includes ten subjects. But it is difficult to recruit ten clinicians or people with M.D. degrees, we decided to recruit at least two participants with M.D. degree, and the rest should have B.S. or B.A. degrees or even higher education background. The sampling method is convenience sampling. I recruit participants by
sending requests to the people I am able to reach and have background in medicine study or related fields, then ask them to recommend people they can reach out to me.

3.2.2 Questionnaire

I send out a questionnaire with the visualization webpage link (https://opal.ils.unc.edu/~luxu/ehr/interface.html) to ask participants about the general view of the visualization interface and to see how the users feel about it and how efficient it is when using the visualization results to make decisions. I inserted some signals in the dataset to see whether the participants can find out them.

There are two parts in the questionnaire. The first half is for all participants, and the second half is for participants with M.D. degree only. For each part, there are 2 to 4 comparatively easy task-based questions tasks (for example: which value do you think is abnormal?), followed by survey questions about how participants think about this visualization.

3.2.3 Semi – structured Interview

The questionnaire about users’ understanding of the visualizations is not sufficient for this user study, thus a semi-structured interview follows up to get more detailed feedback for the visualizations. According to participant’s answers of questionnaire, I asked questions such as why did they choose this answer, and what do they think about this visualization interface and why.
4. Limitations

To engage more real users, the visualization interface should be tested by the real user and be tested with complex, large and real dataset.

The first limitation of this project is that it used small-size simulated dataset. There are potentials that some important signals and patterns of the health-related information are missing in this dataset, so that we cannot evaluate whether the visualization interface works for such signals and patterns. My approach to use the normal range of measures to generate artificial data, then add outliers and trend according to the samples can make the dataset more realistic, but still it cannot be compared with the real, large-size and complex dataset.

The second limitation is using convenient sampling method to recruit participants for user study. The participants for the visualization interface tasks, surveys and user feedback are mostly made up with the Master Students and the Ph.D. candidates who are real users and are not fully representative for the entire user group.
SYSTEM DESIGN

The visualization of the system interface is finished in JavaScript, HTML, CSS programming languages. The JavaScript Libraries that applied to the project include D3, Simple Statistics, jQuery and Bootstrap.

1. User Requirement

   a. The visualization interface should be easy to understand and require minimum efforts to learn how to use and interpret it.

   b. The visualizations of this project are clear and simple, without very fancy animations.

   c. The visualization interface should display all the measures mentioned above in the data collection part.

   d. It should be easy to find the outliers and abnormal trend.

2. System Functions

   The table of the interface has the measures taken from each clinic visit and are displayed in the column corresponding to the date of visit. In between the two clinical visit columns are the scatter plots of self-reported data measured between these two doctor visits.
Table: Measures and Data Collection

<table>
<thead>
<tr>
<th>Date</th>
<th>Weight (lb)</th>
<th>Blood Pressure (mm Hg)</th>
<th>Pulse (BPM)</th>
<th>Pulse Oxygen</th>
<th>AIC</th>
<th>Glucose (mg/dL)</th>
<th>Medicine</th>
</tr>
</thead>
<tbody>
<tr>
<td>11/20/17</td>
<td>129</td>
<td>70 / 100</td>
<td>78</td>
<td>82</td>
<td>0.8</td>
<td>130</td>
<td>med a</td>
</tr>
<tr>
<td>12/18</td>
<td>131</td>
<td>70 / 100</td>
<td>78</td>
<td>76</td>
<td>0.9</td>
<td>140</td>
<td>med a</td>
</tr>
<tr>
<td>2/4/18</td>
<td>130</td>
<td>74 / 104</td>
<td>76</td>
<td>83</td>
<td>1.1</td>
<td>140</td>
<td>med b</td>
</tr>
<tr>
<td>3/4/18</td>
<td>130</td>
<td>74 / 104</td>
<td>76</td>
<td>83</td>
<td>1.1</td>
<td>140</td>
<td>med b</td>
</tr>
<tr>
<td>5/17/18</td>
<td>130</td>
<td>70 / 100</td>
<td>77</td>
<td>79</td>
<td></td>
<td>160</td>
<td>med c</td>
</tr>
</tbody>
</table>

Figure 1 Visualization Interface

2.1 Measure Types

The measures in the table can be selected by the doctors based on the patient type. For this visualization project particularly for diabetes patients, the measures selected are mentioned in the data collection part. Among them are the measures of vital signs such as weight, Blood Pressure, Pulse and measures, and for the diabetes patients specifically, Blood Glucose. The measures displayed in the interface meet requirement c.

2.2 Measure Display

The frequency of user-reported data is more frequent than clinical reported data. Thus, between each two measures which recorded in the clinic/hospital, there are a scatter plot
of the data measured by the patient themselves or the equipment they wear. This function applied two-dimensional visualization, which is easy to interpret for users and satisfies the requirements a and b.

<table>
<thead>
<tr>
<th></th>
<th>11/20/17 17:00</th>
<th>1/2/18 14:04</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>129</td>
<td>131</td>
</tr>
</tbody>
</table>

*Figure 2 Self-reported data and clinical-reported data*

2.3 Self-reported Measures

The scatter plots start and end with the clinic-reported data before and after that period. Each circle is a data point that the patient record.

For the measures with upper and lower limits, such as Blood Pressure, Pulse, Pulse Oxygen, AIC and Glucose, there are blue dashed lines in the chart for the limits. The outliers beyond the upper or lower limits are in red color.

*Figure 3 Self-reported data*
The line in gray is a smoothing line connecting all the circles to emphasize how the trend of the measure changes.

Some measures are single value, but Blood Pressure has compounded values (two values), which include systolic blood pressure and diastolic blood pressure. The measure with two values is displayed as scatter plots with same y-axis scales.

![Figure 4 Measure of two values](image)

Users can mouse over the circle to read more details, containing the exact value and date time.

![Figure 5 Mouseover function](image)

The functions mentioned above meet the requirements a and d. Thus, all aspects of the function design meet the four user requirements.
USER STUDY AND ANALYSIS

The user study is the evaluation part for the visualization interface of the information system. The user study includes questionnaire of task-based questions and survey, and semi-structured interview for each participant. I take around five minutes to walk through the visualization interface, introducing the participant how each of the function works. Then they are asked to answer the questionnaire made up of single-answer-selection questions and open-ended questions. The process of finding answer for each question is similar to the process that the doctors look into the patient’s data and find out more insights. Afterwards I interview them about how did they answer the questionnaire and what are their opinions for the visualizations. In general, it takes around 20 min to finish all the user study parts. The recruit letter, questionnaire and semi-structured interview questions are in the appendix.

1. Participants

Ten participants were recruited for the questionnaires and semi-structured interviews. Among them, two are M.D. or M.D. Candidates, four are Computer Science Ph.D. Candidates, one is Library and Information Science Ph.D. Candidate, one is Ph.D. Candidate in Phycology, one is Ph.D. Candidate in Education, and one is Master of the School of Government.
2. Analysis of Evaluation Results

Q1. Find out the Glucose value of this patient on the day of November 28, 2017.
All participants found the correct answer, which indicates that it is not difficult to learn and find values from the scatter plot.

Q2. Is this Glucose value similar to the Glucose value in about three months, February 10, 2018?
Only 30% of participants selected the correct answer for this question. In the follow-up interviews, I found it was mostly because the participants have different understanding of “similar to”. This question should be more specific and provide explanation of what is the meaning of “similar to”.

Q3. Is the trend of Glucose values during the last two clinical visits (March 4, 2018 and May 17, 2018) similar to the trend of pulse values during the same time period?
All participants found the correct answer, which indicates that trend of the scatter plots is not difficult to figure out.

Q4. Does the patient self-reported Glucose data match the readings from doctor’s office during December 30, 2017 to February 2, 2018?
All participants found the correct answer, which indicates that comparison of clinic reported data and patient self-reported data are obvious.

Q5. Was it easy to find values for a specific date?
Most participants (90%) think it is not difficult to find specific values.

3. Limitations and Future Work

According to the question 7 (What do you find most difficult or limited about this visualization?) and the user study part, the limitations are as follows:

- The scatter plots are overlapping each other and hard to read if the time interval in between is very small.
- Users are not able to read all the measures at one time. Instead, they have to hover each plot to read the value.
- The scale of y-axis is based on the observed maximum and minimum value, thus the variance of different measures cannot be compared.
Future work can be focused on how to solve the overlapping problems if there are too many values during a short period time. Also, the moved lines can be added so that it will be easier to align the values and compare with each other.
CONCLUSION

The research question for this project is how to visualize health related measures of diabetes patients, combined both patient self-reported and clinical-reported information, to assist with clinicians’ understanding of these data and help them to make decisions.

The doctor was interviewed to collect user requirements for the visualization design. The dataset for visualization design and user study is artificial, because the sensitivity of patient clinical data and the absence of user self-reported data in current EHR systems.

After design and development of visualization interface, the user study including task-based questions, surveys and semi-structured interview was conducted to test the efficiency, effectiveness and satisfaction of the visualization interface.

Overall, the visualization interface of EHR System satisfies the requirements of the doctors. The way it combines the clinic-reported data and patient self-reported data is clear and easy to learn. But there are still some functions need to be added to make it work better.
REFERENCE


services and on mortality in the working-age population”. UKK Institute for Health Promotion Research, Tampere, Finland. May 1997 Volume 50, Issue 5, Pages 517–528.


APPENDIX

Questionnaire

1. Find out the Glucose value of this patient on the day of November 28, 2017.
   
   ___

2. Is this Glucose value similar to the Glucose value in about three months, February 10, 2018?
   
   Yes  No

3. Is the trend of Glucose values during the last two clinical visits (March 4, 2018 and May 17, 2018) similar to the trend of pulse values during the same time period?
   
   Yes  No

4. Does the patient self-reported Glucose data match the readings from doctor’s office during December 30, 2017 to February 2, 2018?
   
   Yes  No

5. Was it easy to find values for a specific date?

   Complete not  Probably not  Not sure  likely  Of course

6. What do you like most about this visualization?
   
   __________________________________________________________

7. What do you find most difficult or limited about this visualization?
   
   __________________________________________________________
The following questions are for participants with M.D. only.

8. Can you find out any outliers and unusual patterns of the values before Jan 4, 2018 which may indicate that the patient’s blood glucose will rise on February 3, 2018?

<table>
<thead>
<tr>
<th>Completely not</th>
<th>Probably not</th>
<th>Not sure</th>
<th>likely</th>
<th>Of course</th>
</tr>
</thead>
</table>

9. After the Medicine started form January 2, 2018, do you think the patient’s glucose was under control?

<table>
<thead>
<tr>
<th>Completely not</th>
<th>Probably not</th>
<th>Not sure</th>
<th>likely</th>
<th>Of course</th>
</tr>
</thead>
</table>

10. Would you find this visualization useful if it becomes part of the Epic systems?

________________________________________________________________________

11. When would you use this visualization?

________________________________________________________________________

12. What would you change this visualization?

________________________________________________________________________
**Recruit letter**

Hello,

I am doing a research project on flowsheet visualizations to help clinicians have a better understanding of diabetes’ data. Can you join me walk through the visualizations and then conduct a questionnaire and semi-structured interview, which will in total cost around 20 min?

Thank you so much. Your effort will be greatly appreciated!

Best regards,
Lu