AMERICAN UNIVERSITY OF BEIRUT

METHODS TO EVALUATE AND REDUCE DATA OVERLOAD IN ELECTRONIC MEDICAL RECORDS

by
MAHER BASSAM AL GHALAYINI

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Engineering
to the Department of Industrial Engineering and Management
of the Faculty of Engineering and Architecture
at the American University of Beirut

Beirut, Lebanon
April 2017

AMERICAN UNIVERSITY OF BEIRUT

METHODS TO EVALUATE AND REDUCE DATA OVERLOAD IN ELECTRONIC MEDICAL RECORDS

by
MAHER BASSAM AL GHALAYINI

Approved by:

Dr. Nadine Marie Moacdieh, Assistant Professor Department of Industrial Engineering and Management	Advisor
Dr. Saif Al-Qaisi, Assistant Professor Department of Industrial Engineering and Management	Member of Committee
Dr. Joumana Antoun, Assistant Professor Department of Family Medicine	Member of Committee

Date of thesis defense: April 27, 2017

AMERICAN UNIVERSITY OF BEIRUT

THESIS, DISSERTATION, PROJECT RELEASE FORM

Student Name: ____

Last   First   Middle

○ Master's Thesis   ○ Master's Project   ○ Doctoral Dissertation

☐ I authorize the American University of Beirut to: (a) reproduce hard or electronic copies of my thesis, dissertation, or project; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes.

☐ I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of it; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes after: **One ---- year from the date of submission of my thesis, dissertation, or project.**
Two ---- years from the date of submission of my thesis, dissertation, or project.
Three ---- years from the date of submission of my thesis, dissertation, or project.

____

Signature   Date

ACKNOWLEDGEMENTS

Working on this research was an exceptional learning experience to me. A journey that was made possible only by the support of and guidance of my distinguished professors and beloved family and friends.

First and foremost, I would like to express my deepest gratitude to my thesis advisor, Dr. Nadine Marie Moacdieh, for her outstanding guidance and support throughout these two years. Her patience, encouragement, motivation, enthusiasm, precious feedback, and immense knowledge have helped me in all the time of research and writing of this thesis. Her office door was always open for any question, ready to help settling any issue, and guide me to the right direction whenever needed.

I would like to thank also the members of my thesis committee, Dr. Saif Al-Qaisi, and Dr. Joumana Antoun for their time, support, and valuable feedback during the course of this thesis. I would also like to thank Dr. Antoun for her assistance in developing medical scenarios and for granting me access to the Family Medicine environment at AUBMC.

I would like to thank the medical personnel at the FM department for their help and participation in the related experiments.

My time at AUB was truly enjoyable, especially in my graduate studies, largely because of my friends whom I share my memories with. Special consideration for the ones who always backed me up in difficult times and encouraged me to keep moving on.

Last but not least, I would like to thank my precious family for their constant love, support, and prayers. Their love, encouragement, and motivation guided me through all my pursuits. My family has been with me every step of the way and this journey would not have been possible without them.

AN ABSTRACT OF THE THESIS OF

Maher Bassam Al Ghalayini for Master of Engineering
Major: Engineering Management

Title: Methods to Evaluate and Reduce Data Overload in Electronic Medical Records

Electronic medical records (EMRs) provide physicians with several important functions that help improve hospital operations. However, the introduction of EMRs has also resulted in unforeseen delays in operations and new types of medical errors. These problems have been found to largely stem from the poor design of the EMR interface, and in particular from the issue of data overload. Nevertheless, it is still unclear how best to design and evaluate an EMR in order to minimize data overload and optimize physician performance. To this end, the goal of this research study is to compare the benefits and limitations of a range of interface evaluation methods for their ability to reflect the performance of physicians using an EMR. In addition, these metrics should be able to pinpoint usability issues in EMRs. The application domain for this research will be the AUB Department of Family Medicine. This research will contribute to the literature on interface evaluation and will help usability professionals develop EMRs that maximize safety and efficiency.

CONTENTS

Appendix

ILLUSTRATIONS

TABLES

DEDICATION

To mom and dad

CHAPTER I

INTRODUCTION

Data overload, or the presence of high data density in an interface, is a problem that negatively affects operators in a variety of domains, such as aviation (Alexander, Stelzer, Kim, & Kaber, 2008), driving (Yang et al., 2013), medicine (Singh, Spitzmueller, Petersen, Sawhney, & Sittig, 2013), graphic design (Grahame, Laberge, & Scialfa, 2004), and cartography (Ruth Rosenholtz, Li, & Nakano, 2007). The concern about data overload stems from its negative effects on attention and performance (Wickens & Schons, 1993); namely, slowing down visual search (Beck, Lohrenz, & Trafton, 2010). In particular, data overload in the medical domain can have severe consequences on patient safety and well-being (Singh et al., 2013). The most prominent example of data overload in medicine is in the Electronic Medical Records (EMRs) – i.e., digital medical charts – used by physicians (Farley et al., 2013).

However, to date, researchers have not been able to determine the best way of detecting and measuring data overload in the medical domain. Several methods and metrics have been proposed in the literature, such as subjective assessment (Kim et al., 2012; Schraagen, Chipman, & Shalin, 2000) and image processing (e.g., Berg, Cornelissen, & Roerdink, 2009; Rosenholtz et al., 2007). Nevertheless, it is still unclear which of these measures is best suited to the medical domain. On the other hand, there are alternative metrics whose potential for measuring data overload has not yet been fully assessed, such as eye tracking.

Thus, the overall goal of this research is to systematically and critically compare the benefits and limitations of different metrics to measure data overload in EMRs. These metrics will be evaluated based on their ability to reflect physician performance while using EMRs and pinpoint usability (i.e., ease of use) problems in EMRs. The application domain for this research will be the AUB Medical Center (AUBMC) Department of Family Medicine (FM), who have developed an in-house EMR that they use for patient assessment. In addition, one of the physicians in the FM Department, Dr. Joumana Antoun, has developed another version of this EMR, where elements that had been on separate pages have been combined in one page. While this change could be highly-useful in terms of reduced page-switching time, the concern is that the effects of data overload might offset these benefits. The presence of these two systems provides a unique opportunity to detect the presence of data overload on what is essentially the same system. The hypothesis is that data overload will lead to longer response times to different tasks and this will be reflected by various metrics. From a human factors perspective, the ground truth will be considered to be the response time; in other words, a display will be considered to have data overload if it leads to longer response time for the same task, all else being equal (Alexander et al., 2008).

The specific aims of this research will then be to 1) outline all the functionalities of the EMR in the FM Department, 2) understand the user needs of physicians in the FM Department with respect to the EMR; and 3) conduct a controlled experiment to gather and evaluate the different suggested metrics in order to identify the most appropriate ones for evaluating data overload. This work is expected to result in improved guidelines for the evaluation and assessment of data overload in EMRs. In turn, this will allow for increased efficiency and, more importantly, safety at AUB and other medical environments.

CHAPTER II

BACKGROUND

A. EMRs

EMRs are considered digital equivalents of patient medical charts and store a wide variety of medical information, including physicians' notes, medical reports, laboratory test results, family medical history, past medication, and treatment (Zakaria & Ghani, 2013). EMRs can thus help make medical operations faster and more efficient, which benefits medical providers and patients alike.

However, current EMRs have failed to fully support this goal. A recent study estimated the number of deaths related to preventable medical errors at 440,000 per year (James, 2013), up from the previous estimate of 98,000 in 1999 (Kohm, Corrigan, & Donaldson, 2000). Increasingly, there is a growing realization that errors involving EMRs are not due to a lack of training or negligence on the part of the doctors, but rather stem from poor EMR display design (Karsh, Weinger, Abbott, & Wears, 2010). Design-related issues can thus present a considerable threat to the efficient operation of hospitals and, more importantly, to the health and well-being of patients. The term poor usability is largely considered the result of poor display design. Several factors can contribute to the poor usability of EMRs, such as poor display feedback (Tasa, Ozcan, Yantac, & Unluer, 2008) and faulty input mechanisms (Clarke, Belden, & Kim, 2015), as well as the problem of data overload (Zakaria & Ghani, 2013), which will be the focus of this research.

B. Data Overload

Data overload can be defined as the presence of a high density of poorly-organized data that leads to negative performance decrements (Moacdieh & Sarter, 2012). Also referred to as display clutter, data overload is a problem that affects users in many complex domains, such as aviation (Alexander et al., 2008) and medicine (M. S. Kim et al., 2012). In the medical domain, data overload has figured most prominently in the EMRs used by physicians (e.g., Farley et al., 2013; Karsh et al., 2010; Singh et al., 2013; Van Vleck, Stein, Stetson, & Johnson, 2007). EMR data overload can take on many forms, such as a high density of alerts (Singh et al., 2013) or simply a large amount of poorly-organized, irrelevant medical information (Bobillo, Delgado, & Gomez-Romero, 2008; Hammond, Efthimiadis, & Laundry, 2011; Weir et al., 2007; Zeng, Cimino, & Zou, 2002; Zhang, Pakhomov, McInnes, & Melton, 2011).

C. Negative Effects of Data Overload

The concern about data overload stems from the reported negative effects it has on performance (Bravo & Farid, 2006; Moacdieh & Sarter, 2015; Yeh, Merlo, Wickens, & Brandenburg, 2003). Data overload has been proven to degrade the detection of change in an interface (Moacdieh & Sarter, 2014), increase memory load (Westerbeek & Maes, 2004), lead to confusion (Ewing, Woodruff, & Vickers, 2006), instill confidence in wrong judgments (Baldassi, Megna, & Burr, 2006), delay visual search (Duftschmid et al., 2013; Henderson & Smith, 2009; Murphy, Reis, Sittig, & Singh, 2012; Ruth Rosenholtz et al., 2007), negatively affect object recognition (Bravo & Farid, 2006), and decrease situation awareness (S. Kim & Kaber, 2009). Other negative effects of data overload include committing errors when extracting information (Zeng et al., 2002) and the presence of a higher likelihood of missing data (Singh et al., 2013). Data overload can also increase mental workload (Yang et al., 2012).

When it comes to the medical domain, and in particular to EMRs, data overload has been shown to prevent physicians from quickly and accurately extracting EMR information, which can compromise both efficiency and safety in the hospital (e.g., Moacdieh & Sarter, 2012; Singh et al., 2013; Wu, Zhu, Cao, & Li, 2015). For example, research has shown data overload to lead to increased EMR use time and difficulty finding relevant information (Duftschmid et al., 2013; Murphy et al., 2012). In addition, EMR data overload can lead to errors when extracting medical information (Zeng et al., 2002), physicians reporting higher workload ratings (Ahmed, Chandra, Herasevich, Gajie, & Pickering , 2011), an inability to obtain the "big picture" of a patient's medical condition (Tully et al., 2013), and difficulty identifying information among noise (Van Vleck et al., 2007).

D. Methods for Assessing Data Overload

Despite the negative effects of data overload in EMRs, it is still unclear how best to evaluate data overload in a given display or set of displays. Several different approaches have been used, the most prominent of which rely on subjective input and feedback from physicians. In one study, for example, a survey was used to determine whether practitioners thought there were too many medical alerts in the EMR (Singh et al., 2013). In other cases, interviews with physicians were carried out to obtain their insight on data overload (Van Vleck et al., 2007). In addition to the more prominent subjective methods, researchers have also occasionally used other techniques. Image processing algorithms (pixel-based algorithms that provide a measure of display usability) have sometimes been used, such as the redundancy measure of Zhang et al. (2011). Only rarely have researchers gathered performance measures such as response time and error rate as a measure of display usability (e.g., Duftschmid et al., 2013; Murphy et al., 2012; Zeng et al., 2002).

Despite this attention to EMR usability, there are significant gaps and limitations in the existing literature. The opinions of users, while certainly beneficial and informative, can often be biased and misguided (Andre & Wickens, 1995). Rather, from a user-centered or human factors perspective, what matters most when evaluating or rating interfaces is the performance of users. However, relatively few studies have focused on performance measures. At the same time, performance measures alone are vague and can only suggest whether a display is potentially useful, but not where the problems within the display are. In addition, performance metrics need to be calculated through an experiment, whereas other metrics, such as image processing algorithms, may be calculated without the need for an experiment to be set up. However, there has not been, to my knowledge, a systematic evaluation of which other measures and approaches are most suitable evaluating EMRs. There is a need to determine which measures correlate best to performance and, crucially, can provide additional information about the problems within a display. These measures will then form a framework for the systematic and objective evaluation of EMRs. The proposed measures for evaluating data overload that I will investigate in this research are eye tracking, mouse clicks, image processing algorithms, and GOMS (Goals, Operators, Methods, and Selection Rules). These can then also be used to create a model of data overload in EMRs that is based on multiple factors.

1. *Eye Tracking*

Eye tracking is a non-invasive, infrared-based technique that is used to trace where a person is looking on a screen (Poole & Ball, 2006). Eye tracking has several advantages as a display evaluation technique. In particular, it is a non-invasive and objective measure (Ellis, 2009). Eye tracking also it makes it possible to trace changing information access strategies over time at a fine-grained level of analysis (Zelinsky & Sheinberg, 1997; Zelinsky, 2008). It can also be used for real time processing, which could form the basis of an adaptive display (Sills, 2015). Moacdieh and Sarter were the first to use eye tracking to systematically evaluate the effects of EMR data overload on physician performance while making diagnoses (2015). They suggested a number of eye tracking metrics that proved to be very beneficial, such as convex hull area and spatial density (Goldberg & Kotval, 1999). However, it is not certain if these metrics generalize to all EMR displays. These metrics will thus be further tested in this research.

The output from an eye tracker is a series of screen coordinates that allow researchers to assess when and for how long users were looking at screen elements. The raw coordinates are used to determine eye *fixations*, or when a person looks at something for a minimum period of time (Munn, Stefano, & Pelz, 2008), and *saccades*, which are rapid jumps between fixations (Findlay, 2004). Fixations and saccades then form the building blocks for several eye tracking metrics, such as mean fixation duration (Beck et al., 2010), mean saccade length (Goldberg & Kotval, 1999), and nearest neighbor index (Di Nocera, Camilli, & Terenzi, 2007).

The metrics are classified into three types Spread, Directness, and Duration. The lower their values the better off, since they suggest that the user was able to find what he/she is searching for easily. Spread metrics are concerned with how much dispersion the fixations were. They represent how much the user sampled from the display to find the information he/she needs. Directness metrics are concerned with the path users followed to find the information. Finally, Duration metrics represent the time the user spent searching and finding what he needs.

Table 1 contains the full list of the different eye tracking metrics that will be used in this research.

Table 1 List of Eye Tracking Metrics that will be tested

Name	Explanation
Spread Metrics	
Convex Hull Area (pixels^2) (Hegarty, De Leeuw, & Bonura, n.d.)	Minimum convex area which contains the fixation points.
Spatial Density (Cowen, Baht, & Delin, 2002)	Number of grid cells containing gaze points divided by the total number of cells.
Nearest Neighbor Index (Di Nocera et al., 2007)	The ratio between (1) the average of the observed minimum distance between points and (2) the mean random distance expected if the distribution were random.
Directness Metrics	
Scanpath Ratio (/sec) (Moacdieh & Sarter, 2015)	Length of the distance from the center of screen to the target, divided by the sum of all saccades lengths.
Mean Saccade Length (pixels)	Average of all saccades lengths.
Backtrack Rate (/sec) (Goldberg & Kotval, 1999)	Backtrack is defined as an angle between two saccades that is greater than 90^o.
Rate of Transitions (/sec) (Goldberg & Kotval, 1999)	Rate of transitions between equal grid cells.
Duration Metrics	
Mean Fixation Duration (Moacdieh & Sarter, 2012)	Mean duration of all fixations within a defined period

Total Fixation Rate (count/sec)	Total fixations divided by the time to complete task.

Although this is a promising measure for evaluating data overload, it does present some challenges. One main challenge is that the eye-mind hypothesis, which states that where a person is looking is where that person's attention is focused (Just & Carpenter, 1978) may not always hold. Other problems could arise during calibration, as not all eyes can be easily traced. Additionally, for some eye trackers, users have to keep their heads steady which could affect their performance.

2. *Image Processing Techniques*

Image processing techniques based on pixel characteristics have also been used to measure data overload (Berg et al., 2009; Chang, Zhang, Liu, Yang, & Li, 2010; Lohrenz, Trafton, Beck, & Gendron, 2009; Ruth Rosenholtz et al., 2007). In some approaches, the algorithm produces a scalar value that represents the level of overload (Berg et al., 2009). Other algorithms create what is called a data overload or clutter map by outputting a number of scalar values that represent the data overload in different locations in the display (Jansen & Van Kreveld, 1998; Mušicki, Suvorova, Morelande, & Moran, 2005). One example of such a metric is the color-cluster metric that relies on the definition of data overload as a function of color variation (Lohrenz et al., 2009). One of the most cited techniques is that of Rosenholtz, Li and Nakano (2007) who proposed the feature congestion, sub-band entropy, and edge density metrics. This approach proposed will be adopted in this research given its widespread use and validation (e.g., Holy, Jezek, Snajberk, & Brada, 2012; Miniukovich & De Angeli, 2014; Pereira & Castelhano, 2014).

The three image processing technique metrics of Rosenholtz et al. (2007) depend on a clutter map to compute the different scalar measurements to estimate the data overload level. The feature congestion metric is based on the notion that it is hard to add an object that can capture attention, sometimes referred to as salient, to a cluttered display. This metric depends on clutter maps that are used to show the difference of a certain characteristic between one area and another to output a measure of clutter. These maps are first developed for color, orientation, and texture separately. For example, the color clutter map shows the difference in color in the display. Then these maps are merged into one complete clutter map to provide a scalar measure of the information in the display.

The sub-band entropy metric assumes that the more clutter in a display the less amount of redundancy present. This method, similar to feature congestion, creates a clutter map and calculates a scalar measure. This metric is best used for local clutter or clutter around a target (Asher, Tolhurst, Troscianko, & Gilchrist, 2013). Finally, the edge density method depends on the number of edges present in the display. The density of edge pixels is calculated as a percentage of the total number of pixels to get a measure of clutter. The three measurements calculated give an estimation of the data overload present in the display. While image processing measures can be highly useful and will be tested in this research, their applicability in human factors research is limited given that they are not affected in any way by user factors.

3. ***GOMS***

Goals, Operators, Methods, and Selection Rules (GOMS) decomposes the task into goals and sub-goals (Berry et al., 2016) to determine the amount of time needed to complete a task (Saitwal, Feng, Walji, Patel, & Zhang, 2010). Goals are the tasks to be completed, operators are the different actions taken to accomplish the goals, and methods are the operators' sequences that should be followed. Finally, selection rules are responsible for choosing the best method to accomplish the goal (Saitwal et al., 2010).

GOMS is one of the methods used to perform Cognitive Task Analysis (CTA), which is used to study human performance in real world situations (Saitwal et al., 2010). GOMS is a technique that evaluates the interface as a whole to quantify its complexity and efficiency (Saitwal et al., 2010). It predicts the execution time, learning time, and errors of a task (John & Kieras, 1996). The limitations of GOMS is that it is originally developed to represent error-free performances done by an expert (Segall, Kaber, Taekman, & Wright, 2013). As such, GOMS cannot account for performance errors that even expert users may commit (Segall et al., 2013). In general, as the predicted time to complete a task increases in a display, the usability of this display decreases.

CHAPTER III

METHODS

This study will be divided into three main phases, which will address the three specific objectives: 1) outline all the functionalities of the EMR in the FM Department, 2) understand the user needs of physicians in the FM Department with respect to the EMR; and 3) conduct a controlled experiment to gather and evaluate the different suggested metrics in order to identify the most appropriate ones for evaluating data overload.

A. Phase I

Phase 1 tackled the first objective of this research, which is to outline all the functionalities of the EMR. As a first step, I had several meetings with Dr. Jumana Antoun, an Assistant Professor of Clinical Family Medicine, to learn more about the EMR system and how it is used by physicians (see Figure 1 for an example of a page in the EMR). This EMR had originally been created by Dr. Ghassan Hamadeh at the department of FM, with additional modifications along the years by him and Dr. Antoun. Dr. Antoun explained about the different tasks that physicians perform as part of their work, and how the EMR supports these tasks.

In addition, Dr. Antoun showed and illustrated the use of another version of the EMR that she has developed. This version has not been used in the clinic yet. It was adjusted as part of this study and compared to the original design as a means of assessing the proposed metrics. Following is the description of the EMR functionalities along with the explanation on the current EMR and the redesigned one.

1. *EMR Functionalities*

The current EMR being used in the FM Department was developed over several iterations following its first release. The new version developed by Dr. Antoun has only been developed recently and has never been tested yet. However, although the interface is different, it provides the same functionalities as the original EMR.

a. Current EMR

The EMR contains several functions that physicians use to improve their operations, including writing notes using the "Progress Notes" tab, reviewing patients' medical history using the ROS tab. The complete set of tabs can be found in Table 2. In this system, every set of data has its own page to be presented in. As a result, these pages have more spaces and lesser data to be read. Therefore, it was considered to be the low-data display in our last phase of this research. Figure 1 presents the interface of the Progress Notes tab of this system.

Table 2: EMR tabs and their uses

Tab	Usage
Front Desk	Open patient encounter for a new visit.
Triage	Enter current patient's vitals, enter diagnosis, and check if he/she requires special services.
Preventive Services	Review all patient's vitals, check his/her vaccination reports, and check if he/she requires any special services.
ROS	Review and update patient disease history.
Progress Notes	Review and enter notes related to the patient, check if the patient requires special services, check his/her current vitals, and enter diagnoses.

Dx	Review and update the diagnosis of the patient.
Referrals	Print referral requests to specialty physicians outside the FM department.
X-Ray	Request X-rays images.
Lab	Request lab tests.
WebLab	Check the results of previous lab tests.
Sick Leaves	Give official sick leaves for AUB personnel.
Rx	Renew medications.
Med Hx	Review previous medications.
ER visits	Check and review the patient's visits to the emergency room.

However, this EMR does lack some functions that are typically used by physicians. The system lacks the ability to connect to other AUBMC departments, due to the fact that it was designed for the FM department only. Because of this problem, FM physicians cannot use this EMR to get the patients' records from other AUBMC departments. To try and solve this problem, physicians usually connect to another system that contains additional information, like updated surgical history, Emergency Room visits, etc. Additionally, this EMR does not have the ability to keep physicians updated on the availability of medicines. So, physicians usually use a website that shows them what medicines are available in the market and what new releases are currently there.

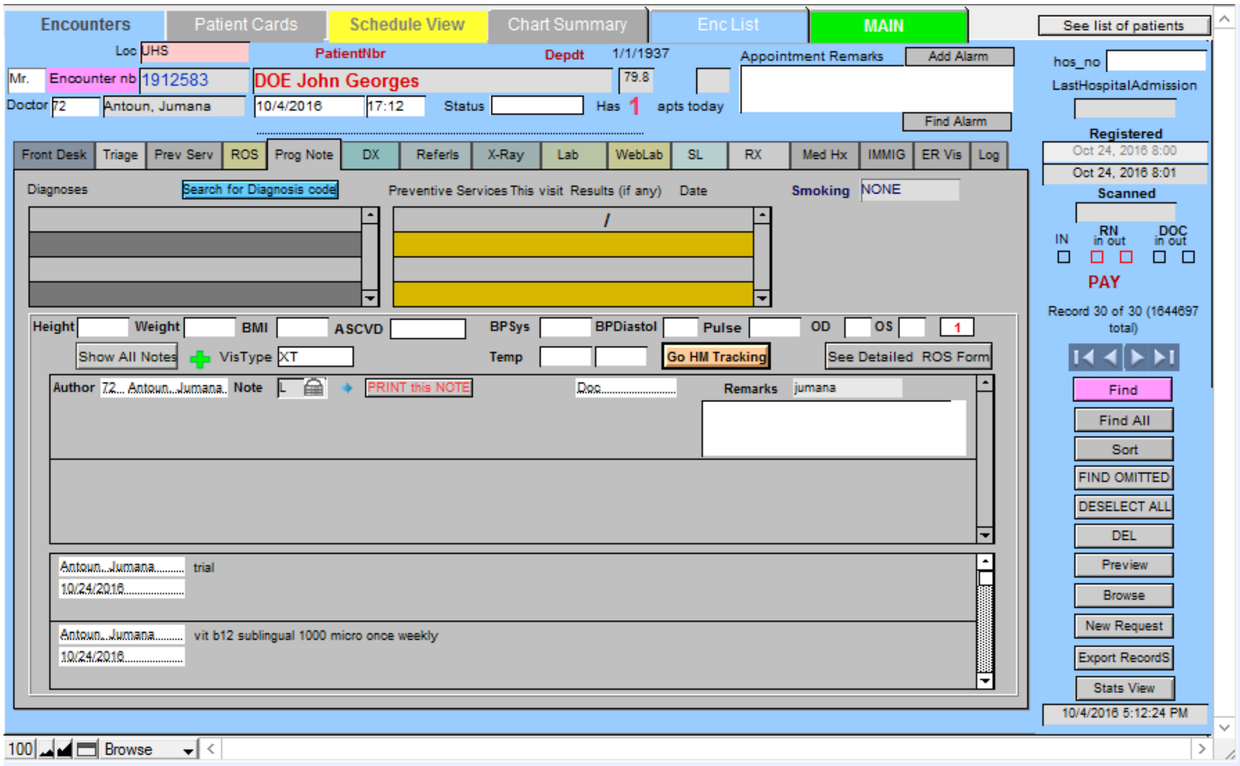


Figure 1. Current EMR interface for the Progress Notes tab

b. <u>Redesigned EMR</u>

The EMR developed by Dr. Antoun contains the data of the three pages ROS, Progress Notes, and Preventives Services in one. This page, which she named Physician, has the same font size, colors, nature, and organization of information as the original EMR, except that the information is stacked together. This has led to an interface that has more data than any one page of the original system. Resulting with a display with high amount of data which risks creating data overload. Figure 2 presents the interface of this new system. Thus, this new EMR page is a good option for the high-data display in the third phase of my research. Therefore, two systems, current and redesigned EMR, will be used in this research to assess the effects of data overload.

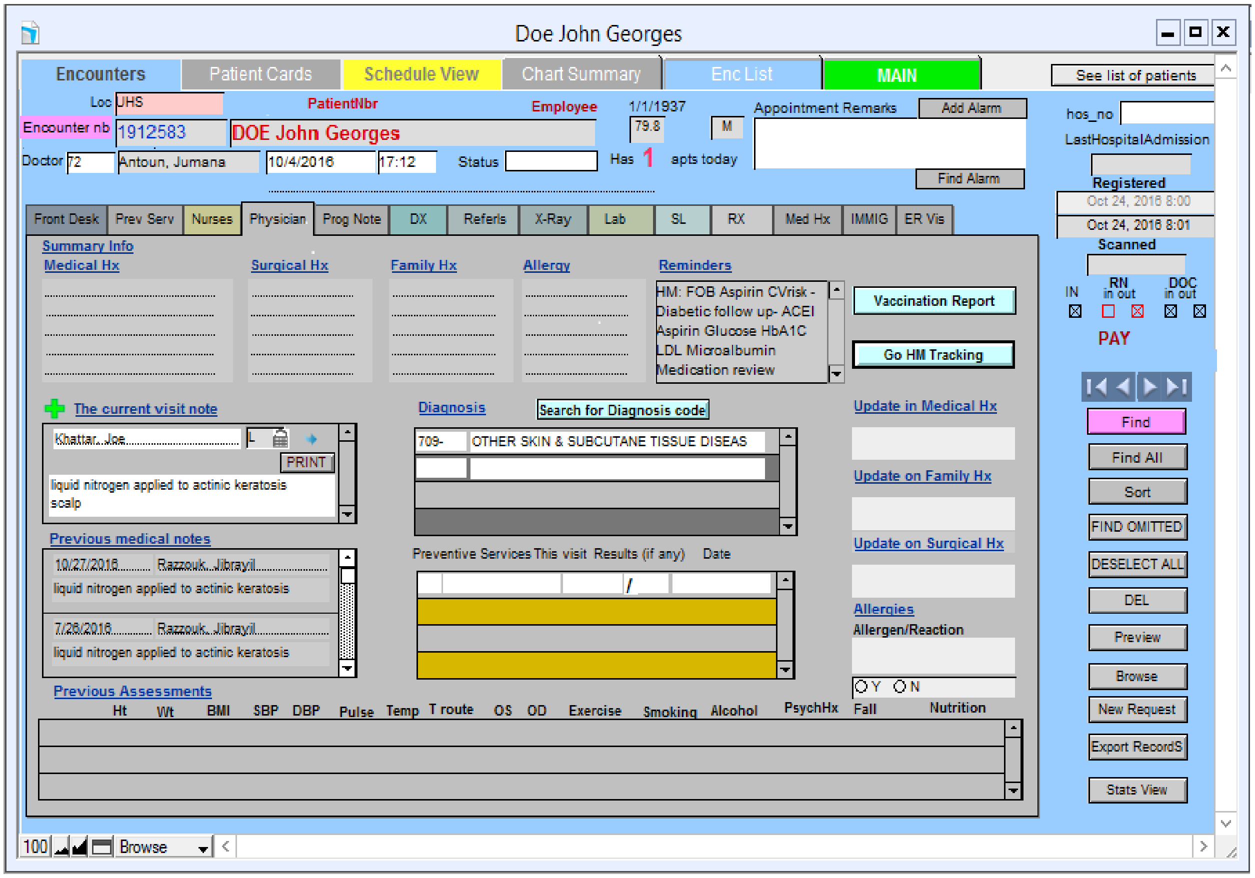


Figure 1. Redesigned EMR interface

a. <u>Cognitive Task Analysis.</u>

Following from my meetings with Dr. Antoun, I conducted a Cognitive Task Analysis (CTA) of the EMR. This is an approach that involves breaking down the use of a system into a series of tasks, and then breaking down each task into its constituent steps. A CTA was performed for each of the displays that formed part of our participant tasks in Phase III (an example of a CTA for one task can be seen in Figure 3). The full results of the CTA can be found in Appendix A.

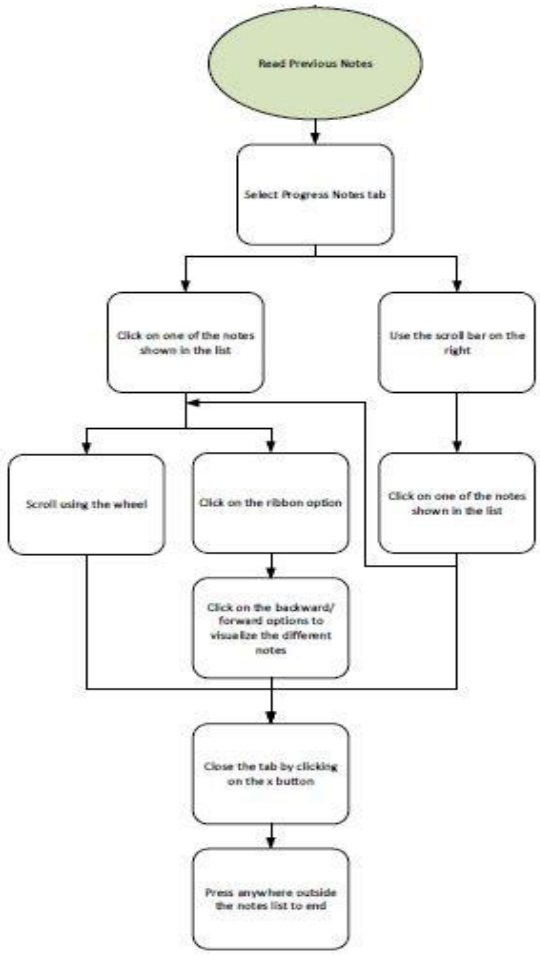


Figure 3. Cognitive Task Analysis for the task of reading previous notes

2. *Discussion*

This introductory phase formed an important first step in this research. This phase provided in-depth insight into the functioning of the EMR, which was necessary in order to be able to carry out Phases II and III. The main conclusions from Phase I are:

a. <u>Wide Range of Functionalities</u>

The current EMR being used provides a wide number of functionalities for FM physicians, ranging from entering notes to reviewing lab information.

b. Set of Steps to Complete Tasks

There are several steps involved in each of the different tasks physicians carry out, involving multiple mouse clicks and/or scrolling.

c. Multiple Ways to Complete Tasks

Due to the multiple design changes and iterations that the EMR has gone through, as well as the need to cater to different physician needs, there are often multiple ways to achieve a given task. For example, physicians can enter the diagnosis using both “Progress Notes” and “Dx” tab, and they can check if any preventive services are needed using both “Preventive Services” and ‘Progress Notes” tabs. This means that physicians often have a choice when it comes to selecting which approach they want to use. This lack of consistency also suggests not all physicians may be using the EMR in an optimal fashion.

B. Phase II

While Phase I provided the in-depth, fundamental knowledge of the system, it was provided by an expert user in Dr. Antoun. What was lacking in Phase I was a better understanding of how the main users of the system, the other physicians and residents in the FM department, actually use the system. Thus, in Phase II, the main goal was to understand the needs of physicians. This includes understanding 1) the typical workflow of FM physicians, 2) what specific tasks physicians perform using their EMR as part of this workflow, and 3) exactly what steps are undertaken using the EMR to achieve these tasks. In order to obtain this information, it was necessary to gather input from several FM physicians to build on what was obtained from Dr. Antoun.

1. *Participants*

The participants in this phase were 5 residents and 1 attending physician, all of whom are currently in the AUB FM Department. An email was sent to all physicians in the FM Department and physicians were asked to respond if they would like to participate in this study. I got in touch to schedule a session at the convenience of the physicians. No monetary compensation was provided to any of the medical personnel. This study was approved by the AUB Institutional Review Board.

2. *Experiment Procedure*

Before meeting with the participants, screen-recording software was installed on their computers by staff at the FM department. Physicians were then asked to record their screens for one clinic session, which typically runs for around four hours and includes a combination of adult and child patient appointments. Each appointment typically lasts for about 20 minutes, including EMR usage time and patient checkup. Physicians were asked to exclude patient names from their screen recordings by only selecting the area that encloses the tabs and the body of the EMR display to be recorded. Any area that showed patient identifiers was thus excluded. Immediately prior to the clinic session, I briefly met with each physician to provide final instructions, solicit any questions/concerns, and provide the consent form for the physician to sign. I was not present during the clinic session. Then, immediately following the completion of the clinic session, I met with the physicians for one hour. The physicians were asked to replay the screen recording and perform a cognitive walkthrough for all of their patient scenarios. In other words, while watching the recording, I asked them to explain all of their different steps and thoughts each time they used the EMR during their clinic session.

For each physician encounter, I then noted what parts of the EMR were used for each task, the sequence in which these tasks were carried out. At the end of the interview, I also asked the physicians for their opinions on what they thought should be changed in every display, what areas they thought contained data overload, and any other comments that they may have. The full questionnaire can be found in Appendix B. The recorded video was destroyed at the end of the meeting.

3. *Results*

Both the performance and subjective results of the six participants were analyzed in order to better understand how physicians and residents currently use their EMRs.

a. Tabs Visited

Table 2 shows an overview of the number of patients seen by each physician during the one clinic session observed. On average, each clinic session took 4 hours and had 6 (SD=2.8) patients. Overall, participants visited 4 tabs per patient on average, where a tab visit consists of any time a participant clicked on a tab to move to that page.

Table 3: Participants and their respective number of patients and tabs visited

Participant	Number of patients	Number of tabs visited	Average number of tabs per patient
1	4	17	4.3

2	9	39	4.3
3	7	23	3.3
4	5	21	4.2
5	3	9	3
6	10	33	3.3
Total	38	142	3.7

As can be seen in Figure 3, participants used almost all of the tabs that are available. However, the tabs that were used not use these tabs in equal numbers. The most visited tabs across all participants were the Progress Notes, ROS, Lab, and Triage tabs with 73, 16, 11, and 10 visits (across all participants), respectively. The complete set of visits is present in Figure 3. The only tabs that no participant used were the X-ray and ER visits tabs.

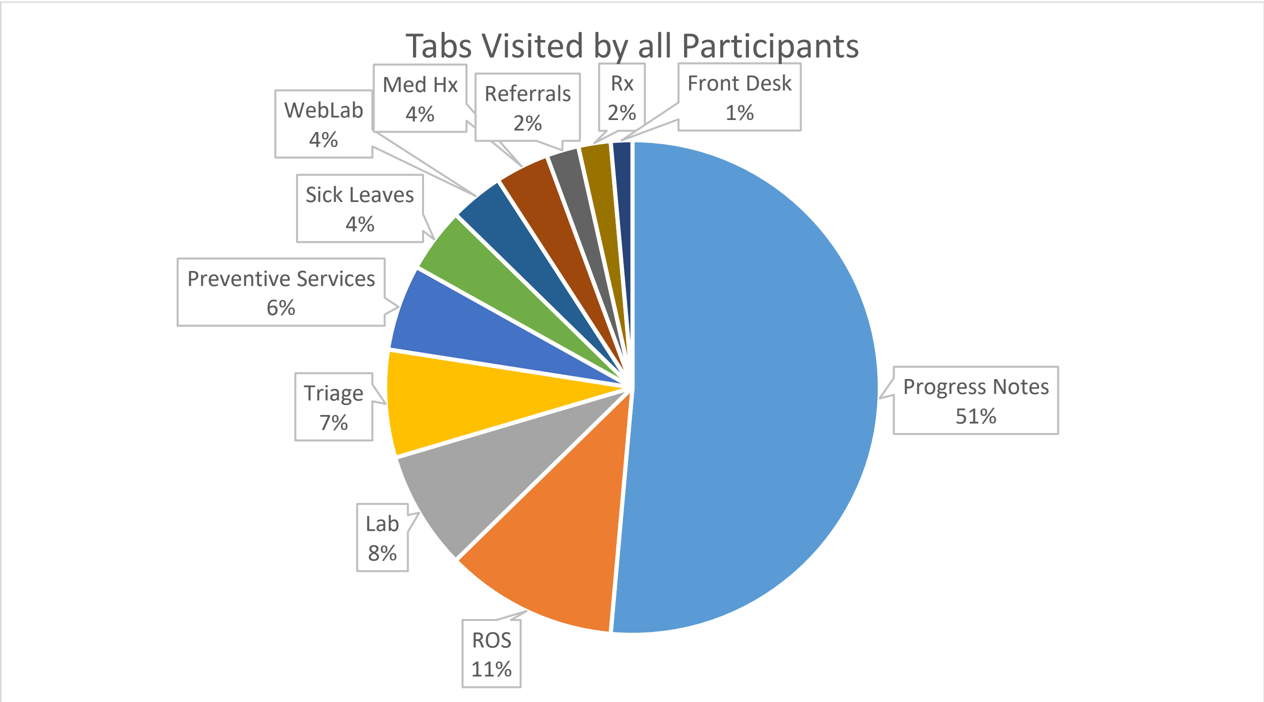


Figure 4: Overall proportion of tabs visited per patient across all participants

In general, the high use of the Progress Notes tab was consistent throughout participants (see Figure 4). Only one participant, Participant 5, visited the Preventive Services equally as much as the Progress Notes tab.

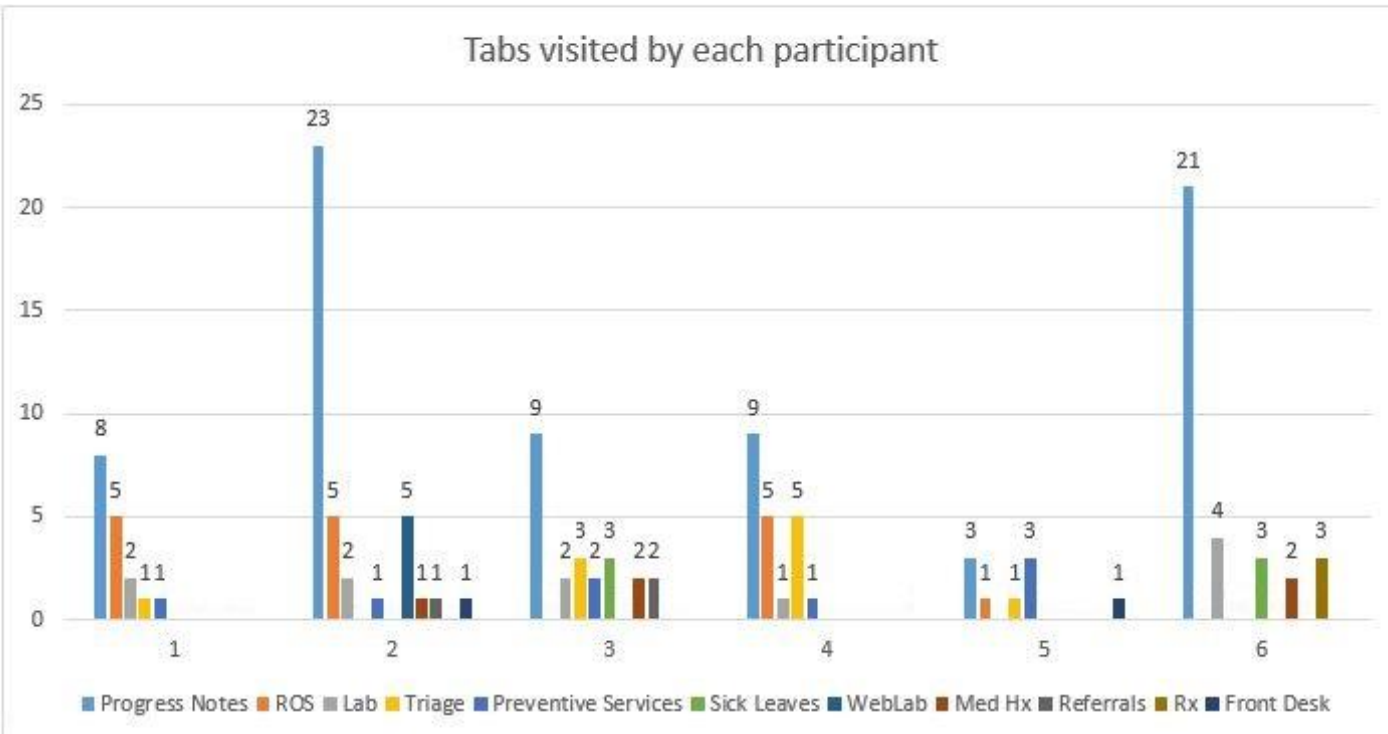


Figure 5. Tabs visited by each participant

b. <u>Sequences</u>

In addition to the tabs that were used, the sequence of these tabs was also of interest. Table 4 shows the general sequence that each participant usually uses during his/her clinic sessions. These sequences are aggregated from all the patient visits for each participant. In other words, for each participant I checked what was the most visited first tab among the patients and placed it as the first one. The same was done for all of the other tabs.

Table 4: Sequences of the Tabs Used by Participants

Participant	General Sequence of Tabs
1	Triage – ROS – Preventive Services – Lab – Progress Notes
2	Progress Notes – ROS – Weblab – Lab – Preventive Services – Front Desk – Referrals – Progress Notes – Front Desk
3	Triage – Progress Notes – Preventive Services – Med Hx – Referrals – Lab – Referrals – Sick Leaves – Progress Notes
4	ROS – Progress Notes – Triage – Lab – Progress Notes
5	Progress Notes – Preventive Services – Progress Notes – ROS – Triage – Front Desk – Preventive Services – Progress Notes - ROS
6	Progress Notes – Lab – Med Hx – Rx – Progress Notes – Sick Leaves

As can be seen in Table 4, half of the participants used the Progress Notes tab as their first tab, and two others used it as their second tab.

c. <u>Levenshtein Distance</u>

Levenshtein Distance is the minimal number of character substitution, deletion, and insertion needed to transform string A to string B. The output of this method is a positive integer with a minimum value of zero. The numerical output is the number of changes done on one string to transform it to the other. Thus, a value of zero means that the two strings are perfectly the same, and the highest value of the output is the length of the longest string. In this study, the sequence of tabs of each participant was transformed into a string input and used to calculate the Levenshtein distance. Each participant's sequence of tabs was transformed into a string of letters, with each letter representing a particular tab (for example the Preventive Services tab was represented by P, the Progress notes tab was represented by N, the ROS tab was represented by R, etc.). In this case, the length of each string was the number of tabs included in the corresponding sequence. The maximum length of all strings was five, which represents the shortest participant sequence, as well as the length of the sequence provided by Dr. Antoun. This was done to make sure that all strings have equal lengths. The goal was to determine the differences in sequence of tabs used between each two physicians as a measure of the degree of similarity between how physicians use their EMR.

The analysis was done using the freely available MATLAB function, strdist, which was downloaded from the MATHWORKS main website. This function takes as input any two strings and provides the Levenshtein distance as output. As can be seen in Table 5, the smallest Levenshtein distance between any two participants' sequences was 3. The average Levenshtein distance across all participants was 4.

To check how closely physicians' sequence of tabs mirrored the one that they were taught, the sequence of each physician was also compared to the sequence provided by Dr. Antoun. This reference sequence consists of Triage, ROS, Progress Notes, Preventive Services, and Lab. The results can be seen in Table 5 also.

Table 5: Levenshtein Distance for the sequences

Participants	**1**	**2**	**3**	**4**	**5**	**6**	**Reference Sequence**
1		3	3	3	5	4	2

2			5	4	4	4	4
3				4	4	4	3
4					5	4	3
5						4	4
6							5

Among the six participants' sequences, the one that are the closest to the reference sequence is that of Participants 1. This sequence only needed two changes to become the same as the reference sequence.

d. Subjective Evaluation

There were two main issues that were highlighted by participants. The first, mentioned by four participants, is that they typically do not have enough time during the clinic session to finish all the required tasks, so patients end up waiting in the waiting area. The second main issue, mentioned by three participants, is that the fonts are not user-friendly and cannot be edited. In other words, the fonts are small and users cannot change the size of words even if the tabs were maximized. As a result, physicians end up taking more time and putting more effort during their work on the EMR. The complete list of issues brought up by participants can be seen in Table 6.

Table 6: Problems faced by the participants when Using the EMR

<table>
<tr><th>Issue</th><th>How many participants</th></tr>
<tr><td>Not enough time to complete all the necessary EMR tasks</td><td>4</td></tr>
<tr><td>The fonts and colors in the WebLab, Progress Notes, Lab, and ROS tabs in particular are not user-friendly</td><td>3</td></tr>
<tr><td>The Triage tab should be the default tab.</td><td>2</td></tr>
<tr><td>More simplified way to edit old ROS data.</td><td rowspan="9">1</td></tr>
<tr><td>Revising the Lab tab for tests arrangement, missing tests, and searching methodology.</td></tr>
<tr><td>Link to Up-to-Date page from the EMR.</td></tr>
<tr><td>Showing the patients who are waiting in a certain place in the EMR.</td></tr>
<tr><td>Slow system.
Old scanned files are not completely present in the EMR.</td></tr>
<tr><td>The EMR is not accessible to other departments.</td></tr>
<tr><td>Entering new notes and editing old ones should be done in more efficient way.</td></tr>
<tr><td>Separating the messages sent to the physicians and replies from the notes.</td></tr>
</table>

4. *Discussion*

The tabs used, their sequence, and the differences between the sequences suggested a number of findings about how physicians use their EMR interface. In particular, the main findings from Phase II of this research were:

a. <u>Importance of the Progress Notes Tab</u>

The results of the different tabs visited during the different patients encounters suggest that participants rely heavily on the Progress Notes tab. They use this tab with every participant, and often more than once. The sequence of tabs shows that it is usually visited early in the patient encounter, as well. Despite the prominence of this page, half of the participants mentioned problems related to its font selection and font color. Given how much it is being used, addressing these issues could significantly impact the experience users have with the display.

b. Consistency Among Participants

What Phase II established clearly is how differently each physician uses his/her EMR display. Despite all participants having been trained in the same way, there were at best three differences between any two physician sequences. With three physicians, there were as many as five differences, for a sequence of length 5, which means that the sequences of these participants were completely different from each other. Moreover, none of the physicians matched the reference sequence that was taught to them by Dr. Antoun, with most physicians showing three, four, and five differences as compared to the reference sequence. This points to a lack of consistency when it comes to the use of the EMR, a fact that has been confirmed in numerous previous studies (e.g. Lanham et al., 2014; Terry, Brown, Bestard Denomme, Thind, & Stewart, 2012). In general, this diversity in the adopted approach is a hallmark of medical practice, where dealing with patients can differ significantly from one physician to another(Terry et al., 2012).

c. Time pressure is an issue.

The number one complaint that participants had about their display was that they do not have enough time to use it properly. This suggests that measures taken to improve EMR displays and reduce operation time would be welcome and necessary.

5. *Conclusion.*

In Phase II, I was able to further understand how physicians work with their EMR during clinics, and what tabs they mostly use. I explored the sequences followed by each participant and inferred that every physician follows his/her own way during clinic sessions, which most of the times was different than the reference one. Additionally, I was able to record the main frustrations and problems faced by the participants when using the EMR.

All of the above were some of the goals of Phase II, but one of the main goals was the selection of the tasks to include in Phase III. Given physicians' limited time, it was important to adopt the tasks that correspond to the tabs that 1) they use the most and 2) have the most problems related to design and information overload. Another important consideration was whether the tab underwent any changes in the new interface designed by Dr. Antoun. Otherwise, comparing across the two interfaces would not result in any differences. For these reasons, we selected the Progress Notes, ROS, and Preventive Services tabs to carry out the tasks of Phase III. Although the above factors also apply to the Lab tab, but it was not changed in the new interface of Dr. Antoun, thus it will not be considered.

C. Phase III

In this third phase, a controlled experiment was conducted with the goal of assessing the proposed data overload evaluation methods. The methods and metrics that were evaluated are: eye tracking, mouse movements and clicks, image processing algorithms, and GOMS. Of these, only the image processing algorithms and GOMS were calculated prior to the start of the experiment since they do not require any participant involvement. The GOMS analysis was done using the CTA that I developed earlier, whereas the algorithms were done through MATLAB using the freely-available algorithms of Rosenholtz et al. (2007). The eye tracking measures and mouse movement and clicks, on the other hand, in addition to the performance measures of response time and error rate, were all collected during the experiment.

1. *Participants*

The participants in this Phase were 13 residents (6 males and 7 females) from the FM Department at AUBMC. Their average age was 29 years (standard deviation (SD) = 2.68) with an average of 2.46 (SD = 1.00) years of experience in the FM Department. Participants had self-reported normal or corrected to normal vision; contact lenses and glasses were allowed. The eye tracking data of one participant was not collected due to problems with the eye tracker software, resulting in a count of 12 residents for the eye tracking results. Participants' self-ratings of proficiency had a mean of 4 (SD = 0.39) on a scale of 1 (poor) to 5 (excellent).

Participants were recruited via email in the same way as in Phase II, and signed an informed consent document at the beginning of the experiment. As an incentive, participants were awarded a $25 gift card to a restaurant, Casper and Gambini's. This study was also approved by the AUB Institutional Review Board.

2. *Experiment Setup*

The location of this experiment was the Cognitive Psychology and Human Computer Interaction Lab of the Psychology Department at AUB (Jessup Hall, Room 107), which contains a Tobii T120 eye tracker. The eye tracker is embedded in the computer screen of size 17 inch with a resolution of 1280 x 1024 pixels. Participants were seated at a distance of about 60 cm. They were asked to provide their answers out loud; for this reason, I sat adjacent to them in order to hear their answers. I also had to be close to the participants in order to tell them what trial to do next, as detailed later.

3. *Experiment Design*

The independent variables were the amount of data (low vs. high) and the type of task (a total of 4 different tasks, as determined from Phase II). The low-data condition refers to the case where relatively little medical data is present on the display. In this study, this corresponds to the original EMR system. On the other hand, the high-data level condition refers to displays that contain a relatively large amount of poorly-organized and irrelevant data, based on the adopted definition of data overload. The high-data display in this study will be the redesigned EMR.

For this experiment, participants had to perform a set of 48 tasks that represent typical patients' cases in the FM Department. Of those 48 tasks, 24 were performed in the case of low-data display (in other words, with the original EMR system), and 24 were performed in the high-data display case (with the new system). There were six trials of each task (i.e., each task was repeated for six different patients' scenarios), for a total of 24 scenarios per experimental condition. Each task took a maximum of two minutes. These scenarios were created by Dr. Antoun, and they were further refined and adjusted after the completion of Phase II of this study. Patients' names and personal information were fictitious but the actual patient data was real (the data were devoid of any names or identifying information). Trials 1, 2, and 3 had the same imagined patient, John Doe, and trials 4, 5, and 6 had the patients John Doe 1, John Doe 2, and John Doe 3 respectively. As a result, the first three trials had the same patient information, and each remaining trial had different information.

The order of presentation of the trials was counterbalanced because the same patient was used for the first three trials and to eliminate the learning effect among trials. To counterbalance the trials, I took the first and the fourth trials together, the second and fifth trials together, and the third and sixth trials together and considered each pair to be a block, after which I counterbalanced the blocks using a Latin Square approach. In this way, the first, second, and third trials never followed each other. The displays were also counterbalanced, such that the odd-numbered participants started with the low-data display, and the even-numbered participants started with the high-data display.

For each trial of each task, physicians were given some background information on the patient that is relevant to the task at hand. The tasks that were used in this experiment are the following:

- Task 1: Review Patient Disease History: In this task, physicians used the ROS display (see Figure B1 in Appendix B). They reviewed patient's disease history and identified whether or not the patient had a given disease/condition within a particular timeframe. An example of this task can be found in Table 9.
- Task 2: Read Previous Notes: In this task, physicians used the Progress Notes display (see Figure 1) of the EMR. They were asked to go through the patient's physician notes and find the most recent note from a given physician.
- Task 3: Check if Preventive Services are Needed: In this task, physicians used the preventive services display (see Figure B2 in Appendix B). They were asked to check the blood pressure, weight, and height of the patients.

- Task 4: Check Vaccination: In this task, physicians used the preventive services display (see Figure B3 in Appendix B). They were asked to check whether a particular vaccine was given to the patients or not.

The first trial of the four tasks is present in Table 7. The complete set of trials and tasks are included in Appendix D. For each trial, there was one patient to do all four tasks on. The name of the patient associated with each trial is also included.

Table 7. Tasks for trial 1

	Tasks for Trial 1: John Doe
1	Before you ask the patient to come into the room, you want to find his chronic diseases. List his problems.
2	Your patient is presenting for dysuria; when you started to prescribe an antibiotic; he reports that few months ago, around May-June he had similar episode and the doctor gave him medication that caused itching. He did not report it to the clinic. He is afraid to take it again. Please find the antibiotics he was prescribed.
3	The patient is worried about high blood pressure today. Check his previous readings in the clinic. Give examples of previous BP readings and their dates.
4	It is winter time and you want to give him flu vaccine and pneumococcal vaccine. He does not recall if he took them ever. Please check his vaccinations status.

The dependent measures in this phase were response time, error rate, the number of mouse clicks recorded while performing the different tasks, and the various eye tracking metrics listed in Table 1. Physicians had to give their answers orally during the experiment and I checked their answers while the experiment was ongoing. The time to complete a task, the number of mouse clicks, and the raw eye tracking data were obtained through the Tobii Studios eye tracking system. The eye tracking metrics of Table 1 were then calculated using MATLAB.

4. *Experiment Procedure*

When the participants arrived to the lab to perform the experiment, they were first given an overview of the experiment and asked to read and sign the consent form. Next, participants were shown the new EMR page and explained how to use it. They were then given the instructions for the experiment and told what the experiment tasks would entail. Each participant was next asked to do a practice scenario with both the original and new EMR pages to become familiar with the experiment process. After this step, the eye tracker was set up and calibrated, which involved asking participants to look at a set of nine points on the screen in order. This first part of the experiment took around 10 minutes. Once calibration was complete, the experiment started. Participants carried out 24 tasks that corresponded to one of the two data-levels conditions. When all the tasks on that one interface were done, participants were given a 5-minute break and then continued the experiment with the other data overload condition. At the end of the experiment, participants were given a questionnaire in which they were asked to rate the amount of data they perceived on each display, as well as providing ratings of mental workload, among other performance assessment questions. This questionnaire can be found in Appendix E.

5. *Results*

The analysis of results for this research involves analysis of the GOMS and image processing algorithm results, both of which were done prior to running the experiments, as well as the subjective, performance, and eye tracking measures of this experiment. Significance was set at p = 0.05.

a. GOMS

The results presented here are based on the results of the CTA in Phase I. KLM-GOMS was used to calculate the times for each task, and the operators used in this model are presented in Table 8 (Card, Moran, & Newell, 1980).

Table 8. CTA GOMS KLM model operators

Step	Time (sec)
K: Key Stroke or button press	0.2
P: Pointing to a target using mouse	1.1
H: Hovering hand over mouse or keyboard	0.4
M: Mentally preparing for doing physical actions	1.35

An example of the analysis that was done for each task of each trial is presented in Table 9 and the full results across each interface and task is shown in Table 10. The complete set of tasks for the two interfaces are presented in Appendix F. As can be seen the time for each task was slightly larger for the high-data interface as compared to the low-data interface. However, given the small sample size and low (in some cases zero) variance, no statistical analysis was performed for these results. The results of these metrics were not correlated with the performance measure and were not tested for difference in the means between the two interfaces since for every task, the trials were very similar to each other, making the variance between each set of results negligible (and sometimes zero).

Table 9. CTA GOMS for task 1 trial 1

Original Interface Task 1 Trial 1		
Search for the "ROS" word	M	1.35
Hover hand to mouse	H	0.4
Drag mouse to "ROS" word	P	1.1
Click using the left mouse button	K	0.2
Look at the medical history	M	1.35
Drag mouse to the first note	P	1.1
Click using the left mouse button	K	0.2
Click using the left mouse button	K	0.2
Read the note	M	1.35
	Total	7.25

Table 10. CTA GOMS times for each trial.

Interface	Trial	Task 1	Task 2	Task 3	Task 4
Original	1	7.25	9.6	8.05	7.05
	2	7.25	9.4	8.45	7.05
	3	7.25	11.6	8.45	7.05
	4	7.25	8.8	7.25	7.05
	5	7.25	8.8	5.75	7.05
	6	7.25	8.8	5.75	7.05
	Average	7.25	9.5	7.2833	7.05
Redesigned	1	7.65	10	8.45	7.45

	2	7.65	9.8	8.85	7.45
	3	7.65	12	8.85	7.45
	4	7.65	9.2	7.65	7.45
	5	7.65	9.2	6.15	7.45
	6	7.65	9.2	6.15	7.45
	Average	7.65	9.9	7.683333	7.45

b. Image Processing

The results of the image processing algorithms are presented in Table 11. For all three metrics of feature congestion, subband entropy, and edge density, the average value for the low-data display was lower than for the high-data display. Looking at the individual pages, the high-data page is higher on feature congestion than two pages, Progress Notes and ROS, of the original interface. On the other hand, the high-data interface has higher subband entropy than all four pages. Edge density for the high-data page is higher than all but one of the four pages of the original interface. As in the results of GOMS, the results of these metrics were not correlated with the performance measure and were not tested for difference in the means between the two interfaces since we only had one measurement for each interface.

Table 11. Image processing values

Interface	Tab	Feature Congestion	Sub-band Entropy	Edge Density
Original	Preventive Services	8.6163	3.9285	0.1256
	Progress Notes	7.4764	3.7015	0.1102

	ROS	7.9533	3.8894	0.1135
	Vaccination	8.785	3.8193	0.1136
	Average	8.207	3.834	0.115
Redesigned	Physician	8.3281	3.9783	0.1173

c. <u>Subjective Metrics</u>

Subjective metrics of amount of data and mental workload were collected for 13 participants using a debriefing post-experiment questionnaire. The scale for both of the ratings was from 1 (poor) to 5 (excellent). The results of this questionnaire were analyzed using a Wilcoxon Exact sign test for ordinal data. This test was selected because it does not assume normal distribution of the data.

Results showed that there was no significant difference in participants' ratings of the amount of data in the two displays ($Z = -586$, $p = 0.558$), with means of 3.84 (Standard deviation (SD) = 0.688) and 3.61 (SD = 1.04) for low-data and high-data displays respectively. On the other hand, results showed that there was a significant difference in the mental workload ($Z = -2.04$, $p = 0.041$) with means of 3.3 (SD = 1.3) and 2.46 (SD = 0.877) for both low-data and high-data displays respectively.

Other than the amount of data and the mental workload required for both displays, participants were asked to point out which display was easier to handle. All except two participants indicated that the high-data display was easier to work with. Of the remaining two, one indicated that the low-data display is easier and the other indicated that they were both the same.

The Pearson correlation of the two subjective measures with the response time for both interfaces are presented in Tables 12.

Table 12. Pearson Correlation between the subjective measures and response time for the both interfaces

Interface	Measure		Task 1	Task 2	Task 3	Task 4
Original	Amount of Data	Correlation	**-0.77**	-0.25	-0.41	**-.72**
		Significance	**0.00**	0.40	0.16	**0.00**
	Mental Workload	Correlation	-0.61	-0.40	-0.34	-.63
		Significance	0.03	0.17	0.25	0.02
Redesigned	Amount of Data	Correlation	**-.74**	**-.70**	**-.73**	**-.73**
		Significance	**0.00**	**0.01**	**0.00**	**0.00**
	Mental Workload	Correlation	-.63	-.65	-.60	-.62
		Significance	0.02	0.01	0.03	0.02

d. Performance Measures

The performance measures in this research were the response time, error rate, and number of mouse clicks needed to complete each task. Error rate was zero as all participants performed their tasks correctly. The other performance measures were averaged across the six trials for each participant and each task.

The performance data was analyzed using a repeated-measures analysis of variance (ANOVA). The assumptions for the ANOVA procedure were tested for each response measure and each task. There were some measures where the normality assumption did not hold, as evidenced using a Shapiro-Wilk test and a normality plot. So, the data of these measures were transformed using either the inverse or logarithmic transform. The metrics for which this was done can be seen in Table 13. Since we needed to compare the means across the interfaces, if one measure violated the assumption, the data of this measure for both displays were transformed.

Table 13. Performance metrics that violated ANOVA's normality assumption

Task	Metric	Transformation
Task 1	Mouse Clicks	Inverse
Task 2	Response Time	Logarithmic
Task 3	Response Time	Logarithmic
	Mouse Clicks	Inverse
Task 4	Response Time	Inverse

Note that since every task has its own display and its own procedure, the differences between measures across the tasks were not compared. ANOVA results for Response Time and Mouse clicks can be found in Tables 14 and 15 respectively.

Table 14. Response time ANOVA results

Response Time (sec)	Original Interface Mean (SD)	Redesigned Interface Mean (SD)	Effect of Task level
Task 1	**6.39 (1.69)**	**4.76 (2.24)**	**F (1,12) = 5.546, p = 0.038** $\boldsymbol{\eta_p^2}$ **= 0.313**
Task 2	24.48 (5.51)	26.37 (10.65)	F (1,12) = 0.292, Not Significant (p = 0.559) $\eta_p^2 = 0.024$

Task 3	1.02 (0.12)	0.97 (0.25)	F (1,12) = 0.923, Not Significant (p = 0.356) η_p^2 = 0.071
Task 4	0.19 (0.23)	0.050 (0.07)	F (1,12) = 2.598, Not Significant (p = 0.133) η_p^2 = 0.178

Table 15. Mouse clicks ANOVA results

Mouse Clicks (count)	Original Interface Mean (SD)	Redesigned Interface Mean (SD)	Effect of Task level
Task 1	**0.40 (0.098)**	**0.69 (0.16)**	**F (1,12) = 9.836, p = 0.009 $\boldsymbol{\eta_p^2 = 0.45}$**
Task 2	**14.78 (4.74)**	**8.85 (4.58)**	**F (1,12) = 10.717, p = 0.007 $\boldsymbol{\eta_p^2 = 0.472}$**
Task 3	0.27 (0.095)	0.35 (0.21)	F (1,12) = 1.875, Not Significant (p = 0.196) $\eta_p^2 = 0.135$
Task 4	**2.68 (0.60)**	**2.26 (0.40)**	**F (1,12) = 7.081, p = 0.021 $\boldsymbol{\eta_p^2 = 0.371}$**

Pearson correlation of the mouse clicks with the response time can be found in Table 16.

Table 16. Pearson Correlation between mouse clicks and response time for both interfaces

Mouse Clicks		Task 1	Task 2	Task 3	Task 4
Original Interface	Correlation	**0.79**	-0.41	-0.32	-0.47
	Significance	**0.00**	0.17	0.29	0.11
Redesigned Interface	Correlation	**.95**	-0.44	-0.21	**-.74**
	Significance	**0.00**	0.14	0.48	**0.00**

e. Eye Tracking Metrics

Eye tracking metrics were collected for 12, rather than 13 participants, due to complete eye tracker device failure for one participant. For the due to tracking and recording problems, for 2 participants, a total of 10 entries were excluded. As with the performance measures, the data was analyzed using a repeated-measures ANOVA.

After averaging the trials for each task and each participant, the assumptions of the ANOVA procedure were checked. In the case of violations of normality, as evidenced using a Shapiro-Wilk test and a normality plot, the data was transformed, as can be seen for specific metrics in Table 17.

Table 17. Eye tracking metrics that violated ANOVA's normality assumption

Task	Metric	Transformation
Task 1	Convex Hull Area Scanpath Ratio	Logarithmic
Task 2	Rate of Transition	Inverse
Task 3	NNI Total Fixations Rate Backtrack Rate Rate of Transition	Inverse
Task 4	NNI Rate of Transition	Inverse

The results of the ANOVA procedure can be seen in Tables 18, 19, 20, and 21.

Table 18. Eye tracking metrics results of task 1

Task 1			
Eye Tracking Metrics	**Original Interface Mean (SD)**	**Redesigned Interface Mean (SD)**	**Effect of Task level**
Directness Metrics			
Backtrack Rate (/sec)	0.0047 (0.00219)	0.005 (0.00271)	F (1,11) = 0.436, Not Significant (p = 0.523) $\eta_p^2 =$ 0.038
Rate of Transitions (/sec)	0.003 (0.00086)	0.0032 (0.00126)	F (1,11) = 0.41, Not Significant (p = 0.535) $\eta_p^2 = 0.036$
Scanpath Ratio (pixels/sec)	1.13 (0.18910)	1.08 (0.23577)	F (1,11) = 0.777, Not Significant (p = 0.397) $\eta_p^2 = 0.066$
Mean Saccade Length (pixels)	63.34 (27.988)	61.76 (26.30)	F (1,11) = 0.062, Not Significant (p = 0.807) $\eta_p^2 = 0.006$
Duration Metrics			
Total Fixation Rate (count/sec)	0.0083 (0.003)	0.0087 (0.0035)	F (1,11) = 0.447, Not Significant (p = 0.504) $\eta_p^2 = 0.042$
Mean Fixation Duration (millisecond)	**126.57 (68.73)**	**96.66 (37.87)**	**F (1,11) = 0.447, p = 0.025** $\boldsymbol{\eta_p^2 = 0.378}$
Location Metrics			
Convex Hull Area (pixels^2)	4.97 (0.25)	4.86 (0.3)	F (1,11) = 1.644, Not Significant (p = 0.226) $\eta_p^2 = 0.13$
Spatial Density	0.064 (0.016)	0.057 (0,023)	F (1,11) = 1.413, Not Significant (p = 0.26) $\eta_p^2 = 0.114$
NNI	0.3 (0.13)	0.24 (0.1)	F (1,11) = 1.457, Not Significant (p = 0.253) $\eta_p^2 = 0.117$

Table 19. Eye tracking metrics results of task 2

Task 2			
Eye Tracking Metrics	**Original Interface Mean (SD)**	**Redesigned Interface Mean (SD)**	**Effect of Task level**
Directness Metrics			
Backtrack Rate (/sec)	0.0049 (0.0023)	0.0053 (0.0029)	F (1,11) = 0.363, Not Significant (p = 0.559) $\eta_p^2 = 0.032$
Rate of Transitions (/sec)	391.45 (97.67)	448.18 (254.17)	F (1,11) = 0.744, Not Significant (p = 0.407) $\eta_p^2 = 0.063$
Scanpath Ratio (pixels/sec)	95.35 (40.70)	100.94 (50.82)	F (1,11) = 0.161, Not Significant (p = 0.696) $\eta_p^2 = 0.014$
Mean Saccade Length (pixels)	**58.37 (18.90)**	**50.75 (13.70)**	**F (1,11) = 5.263, p = 0.042** $\boldsymbol{\eta_p^2 = 0.324}$
Duration Metrics			
Total Fixation Rate (count/sec)	0.008 (0.0032)	0.0084 (0.004)	F (1,11) = 0.308, Not Significant (p = 0.59) $\eta_p^2 = 0.027$
Mean Fixation Duration (millisecond)	114.80 (67.45)	104.64 (58.57)	F (1,11) = 1.035, Not Significant (p = 0.331) $\eta_p^2 = 0.086$
Location Metrics			
Convex Hull Area (pixels^2)	**280855.43 (47459.28)**	**226138.60 (66561.86)**	**F (1,11) = 5.266, p = 0.042** $\boldsymbol{\eta_p^2 = 0.051}$
Spatial Density	0.14 (0.31)	0.13 (0.03)	F (1,11) = 1.186, Not Significant (p = 0.299) $\eta_p^2 = 0.097$
NNI	0.83 (0.34)	0.76 (0.48)	F (1,11) = 0.027, Not Significant (p = 0.873) $\eta_p^2 = 0.002$

Table 20. Eye tracking metrics results of task 3

Task 3			
Eye Tracking Metrics	**Original Interface Mean (SD)**	**Redesigned Interface Mean (SD)**	**Effect of Task level**
		Directness Metrics	
Backtrack Rate (/sec)	261.43 (129.05)	255.42 (109.38)	F (1,11) = 0.059, Not Significant (p = 0.813) $\eta_p^2 = 0.005$
Rate of Transitions (/sec)	348.54 (95.72)	358.61 (101.55)	F (1,11) = 0.115, Not Significant (p = 0.741) $\eta_p^2 = 0.01$
Scanpath Ratio (pixels/sec)	**55.53 (19.85)**	**42.93 (21.98)**	**F (1,11) = 5.058, p = 0.046** $\boldsymbol{\eta_p^2 = 0.315}$
Mean Saccade Length (pixels)	**70.86 (21.24)**	**63.17 (20.13)**	**F (1,11) = 5.887, p = 0.034** $\boldsymbol{\eta_p^2 = 0.349}$
		Duration Metrics	
Total Fixation Rate (count/sec)	138 (51.16)	138.01 (46.12)	F (1,11) = 0.00, Not Significant (p = 1) $\eta_p^2 = 0.00$
Mean Fixation Duration (millisecond)	129.48 (66.27)	107.91 (54.23)	F (1,11) = 5.713, Not Significant (p = 0.906) $\eta_p^2 = 0.342$
		Location Metrics	
Convex Hull Area ($pixels^2$)	221684.02 (40994.41)	196213.71 (45196.38)	F (1,11) = 1.672, Not Significant (p = 0.223) $\eta_p^2 = 0.132$
Spatial Density	0.1073 (0.02)	0.1005 (0.024)	F (1,11) = 0.885, Not Significant (p = 0.367) $\eta_p^2 = 0.074$
NNI	4.15 (1.42)	4.035 (1.41)	F (1,11) = 0.063, Not Significant (p = 0.806) $\eta_p^2 = 0.006$

Table 21. Eye tracking metrics results of task 4

Task 4			
Eye Tracking Metrics	**Original Interface Mean (SD)**	**Redesigned Interface Mean (SD)**	**Effect of Task level**
Directness Metrics			
Backtrack Rate (/sec)	0.0052 (0.0024)	0.0051 (0.0027)	F (1,11) = 0.005, Not Significant (p = 0.946) $\eta_p^2 = 0.001$
Rate of Transitions (/sec)	311.36 (91.38)	300.76 (85.62)	F (1,11) = 0.242, Not Significant (p = 0.632) $\eta_p^2 = 0.022$
Scanpath Ratio (pixels/sec)	39.95 (12.07)	36.86 (10.27)	F (1,11) = 0.465, Not Significant (p = 0.509) $\eta_p^2 = 0.041$
Mean Saccade Length (pixels)	**71.65 (21.96)**	**65.31 (17.63)**	**F (1,11) = 4.744, p = 0.049** $\boldsymbol{\eta_p^2 = 0.301}$
Duration Metrics			
Total Fixation Rate (count/sec)	0.0089 (0.0032)	0.0089 (0.00377)	F (1,11) = 0.0001, Not Significant (p = 0.998) $\eta_p^2 = 0.0001$
Mean Fixation Duration (millisecond)	105.59 (55.01)	104.28 (49.8)	F (1,11) = 0.02, Not Significant (p = 0.889) $\eta_p^2 = 0.002$
Spread Metrics			
Convex Hull Area (pixels^2)	**204511.69 (20855.20)**	**183652.55 (25531.58)**	**F (1,11) = 5.1, p = 0.045** $\boldsymbol{\eta_p^2 = 0.317}$
Spatial Density	0.0085 (0.016)	0.09 (0.01)	F (1,11) = 1.496, Not Significant (p = 0.247) $\eta_p^2 = 0.12$
NNI	4.78 (0.80)	4.45 (1.32)	F (1,11) = 0.681, Not Significant (p = 0.427) $\eta_p^2 = 0.058$

A summary of the eye tracking metrics that showed significant difference between in the ANOVA results across all the tasks can be seen in Table 22.

Table 22. Summary table of significant eye tracking metrics

Eye Tracking Metrics	Original Interface Mean (SD)	Redesigned Interface Mean (SD)	Effect of Task Level
		Task 1	
Mean Fixation Duration (millisecond)	126.57 (68.73)	96.66 (37.87)	F (1,11) = 0.447, p = 0.025 $\eta_p^2 = 0.378$
		Task 2	
Mean Saccade Length (pixels)	58.37 (18.90)	50.75 (13.70)	F (1,11) = 5.263, p = 0.042 $\eta_p^2 = 0.324$
Convex Hull Area (pixels^2)	280855.43 (47459.28)	226138.60 (66561.86)	F (1,11) = 5.266, p = 0.042 $\eta_p^2 = 0.051$
		Task 3	
Scanpath Ratio (pixels/sec)	55.53 (19.85)	42.93 (21.98)	F (1,11) = 5.058, p = 0.046 $\eta_p^2 = 0.315$
Mean Saccade Length (pixels)	70.86 (21.24)	63.17 (20.13)	F (1,11) = 5.887, p = 0.034 $\eta_p^2 = 0.349$
		Task 4	
Mean Saccade Length (pixels)	71.65 (21.96)	65.31 (17.63)	F (1,11) = 4.744, p = 0.049 $\eta_p^2 = 0.301$
Convex Hull Area (pixels^2)	204511.69 (20855.20)	183652.55 (25531.58)	F (1,11) = 5.1, p = 0.045 $\eta_p^2 = 0.317$

The Pearson correlation between the eye tracking metrics and response time are found in Tables 23 and 24.

Table 23. Pearson Correlation between the eye tracking metrics and response time for the original interface

Original Interface		Task 1	Task 2	Task 3	Task 4
Convex Hull Area	Correlation	**-0.72**	-0.63	-.592	-.57
	Significance	**0.01**	0.03	0.04	0.05
Spatial Density	Correlation	-0.65	-0.36	-0.56	**-.75**
	Significance	0.02	0.26	0.06	**0.00**
Nearest Neighbor Index	Correlation	-0.49	0.44	0.36	**.92**
	Significance	0.11	0.15	0.25	**0.00**
Mean Fixation Duration	Correlation	-0.51	-0.69	-0.49	-0.36
	Significance	0.09	0.01	0.10	0.25
Scanpath Length	Correlation	-0.40	0.54	0.36	-0.23
	Significance	0.20	0.07	0.26	0.48
Scanpath Ratio	Correlation	0.50	0.55	0.51	-0.15
	Significance	0.10	0.07	0.09	0.65
Mean Saccade Length	Correlation	-0.57	**-0.72**	**-.72**	-0.52
	Significance	0.05	**0.01**	**0.01**	0.08

Table 24. Pearson Correlation between the eye tracking metrics and response time for the redesigned interface

Redesigned Interface		Task 1	Task 2	Task 3	Task 4
Convex Hull Area	Correlation	**-.73**	**-.70**	**-.83**	-0.56
	Significance	**0.01**	**0.01**	**0.00**	0.06
Spatial Density	Correlation	-0.49	**-.70**	**-.89**	**-.90**
	Significance	0.11	**0.01**	**0.00**	**0.00**
Nearest Neighbor Index	Correlation	-0.55	0.06	.61	0.54
	Significance	0.06	0.86	0.03	0.07
Mean Fixation Duration	Correlation	-0.48	-0.52	-.60	-0.11
	Significance	0.11	0.08	0.04	0.74
Scanpath Length	Correlation	-.69	-0.09	-0.12	-0.38
	Significance	0.01	0.77	0.70	0.23
Scanpath Ratio	Correlation	-0.21	-0.11	-0.12	-0.06
	Significance	0.50	0.73	0.71	0.84
Mean Saccade Length	Correlation	-.68	**-.84**	-.69	**-.60**
	Significance	0.01	**0.00**	0.01	**0.04**

3. *Discussion*

The overall goal of this study was to systematically and critically compare the benefits and limitations of different metrics to measure data overload in EMRs. These metrics will be evaluated based on their ability to reflect physician performance while using EMRs and pinpoint usability (i.e., ease of use) problems in EMRs. The hypothesis was that the higher-data interface will lead to longer response times to different tasks, more inefficiencies in extracting data, more confusion, and less satisfaction overall by physicians, and this will be reflected by various metrics. However, it was expected that the higher-data interface will lead to fewer mouse clicks, which was confirmed in this study for three out of four tasks. This provides evidence that even though the new, high-data display does not reduce response time, it does reduce the steps that physicians have to go through, which was the original aim of the interface.

Table 25 summarizes the main results of Phase III. In general, results did not support our hypothesis on the higher-data interface leading to longer response times. At the same time, the high-data interface did generally lead to fewer mouse clicks, as expected. These results suggest that the additional data in the high-data interface was not significant enough to lead to performance decrements. Instead, the new EMR design with data combined in one location appears to lead to better physician performance. This could be explained by the fact that the redesigned interface, unlike the current, only contained the data that are relevant to the tasks. The current interface contains a lot of irrelevant data that could cause confusion and inefficiency in extracting data from the system. This finding is telling in that it suggests the benefits of combining information in one location can offset any costs of data overload, under the condition that this information are all relevant and usable to the tasks in hand.

Table 25. Phase III main results

Metrics	Reflects response time?	Pinpoints usability problems?
GOMS	No	No
Image processing algorithms (each of the three alone)	No	Yes
Subjective feedback (Rate of Amount of Data)	No	Yes
Subjective feedback (Rate of Mental Workload)	Yes	No
Subjective feedback (Preferences)	No	Yes
Eye tracking data (the ones that showed significance)	Yes	Yes

a. <u>GOMS</u>

The GOMS approach is derived from the Model Human Processor (Card et al., 1980), and is thus grounded in human behavior modeling. Despite this fact, the response time results obtained using KLM-GOMS, one variation of GOMS, were very different from the actual response times obtained, suggesting that this is not a reliable technique to estimate the time it takes to perform an EMR task.

b. <u>Image Processing</u>

The image processing algorithms did not correlate with performance measures. This is understandable, given that these measures are not related to any aspect of the human or human performance. This provides further validation that, from a human factors perspective, these measures are not extremely useful as an indication of the effects of clutter.

However, their benefit lies primarily in their ability to pinpoint potential usability problems. For example, here all three algorithms – feature congestion, subband entropy, and edge density. An increase in feature congestion suggests there are now more variation in the number of features – colors, luminance, etc. – present in the display, which makes sense given the additional items. An increase in subband entropy suggests there is less symmetry in the high-data display, which is also a by-product of the increased number of items. Finally, an increase in edge density is also obvious given that an increase in the number of items will lead to an increase in the number of edges.

The recommendations from these algorithms would be to decrease the variation in colors in the display, make the display more symmetric, and remove a few of the items. Given the performance effects obtained, these changes will likely not lead to any significant improvements in terms of performance. The only improvement might be in aesthetics. However, it is important to note that the algorithm values obtained are relatively high compared to other cluttered images, where 25 distractors in one image only gave a feature congestion measure of around 6.5 (Rosenholtz, Li, Jin, & Mansfield, 2005). In this research, both the low and the high-data displays obtained feature congestion measures above 8. It would appear that both displays are equally high-data. At the same time, physicians are experienced enough that it does not make a difference.

In summary, image processing measures are useful but not able to reflect the human factors associated with medical work. The algorithms cannot reflect the fact that users were not affected by the data load, which is evident in the low correlation values.

c. Subjective Data

This form of data is one of the most common approaches to discovering usability problems with EMRs. In this study as well, participants provided several suggestions as to how the design could be improved. Another means of obtaining subjective data is through ratings of the amount of data, as well as ratings of mental workload. However, as with performance measures, the concern with subjective ratings is that they may not accurately reflect performance. Users are known to be misguided about their own performance (Andre & Wickens, 1995). This concern proved to be well founded, given that the correlation between both sets of subjective ratings (data and mental workload) were never greater than +/- 0.75. This suggests that subjective ratings are not an ideal approach to elicit performance effects, although they are valuable for usability problems. In addition, the subjective ratings of mental workload did indicate that participants find the high-data display easier to use.

d. Eye Tracking

Several eye tracking metrics were significantly different in the high-data display. In particular, convex hull area, mean saccade length, scanpath ratio, and mean fixation duration were significantly affected and they notably all decreased in the high-data display. This indicates that the high-data display was easier to obtain information from; in other words, participants needed to expend fewer attentional resources in order to obtain their needed information.

These metrics also incorporate the three groups of interest and indicate which aspects of the display were problematic. Convex hull area is a spread metric, meaning that it indicates how much of the display the user had to cover. A lower convex hull area here means that there was less of the display that participants had to sample, meaning fewer dispersion of attention. Mean saccade length and scanpath ratio, on the other hand, are directness metrics, meaning that they indicate how efficiently the user arrive at the information they needed. The lower values in this case suggest participants moved in systematic fashion towards the information, and there was little confusion or uncertainty. The smaller mean saccade length means they moved in smaller steps towards the target, usually an indication of more efficient processing. As for the lower scanpath ratio, it indicates that participants could efficiently get to the target they were looking for. Finally, mean fixation duration is a duration metric that indicates how much difficulty users had with extracting data from the display. Once again, it would appear that participants had less difficulty in the high-data case.

Of these metrics, correlation with response time showed that convex hull area and mean saccade length were the ones that were highly correlated with response time, lending further credence to the belief that these metrics reflect the performance effects of data overload. Spatial density also appeared to be promising in this regard, which supports the findings of Moacdieh and Sarter (2015). However, while these eye tracking metrics can indicate how users' attention was affected and the nature of the display problems (e.g., poor discriminability, no guidance to the target, etc.), it is difficult to pinpoint exactly which areas of the display or which features led to these problems. This would require closer inspection of the scanpath.

CHAPTER IV

LIMITATIONS AND FUTURE WORK

There were a number of limitations in this study, chief of which was the low number of participants in both Phase II and Phase III. Moreover, all participants in Phase III were residents, so I did not explore how physicians would have handled with the new EMR. As for the tasks done in this Phase, they only represent a small portion of what physicians really do with the EMR. Moreover, this experiment was done in a very controlled environment. Unlike the experiment, during their clinic sessions, physicians face pressure and encounter many interruptions and distractors which could lower their performance and cognitive ability. Additionally, the eye tracking metrics studied were originally used for searching tasks, in which information are placed in the same display, here they were used even though participants had to change between tabs to get what they want. Finally, the metric “Scanpath Ratio” is usually the ratio of the distance between the user’s first fixation and the target, and scanpath length. This was not the case here in which I assumed that participants will start their search (i.e. first fixation) from the center of the display, so the distance was taken to be from the center of the display to the target for simplicity.

Future work will look to increase the sample size and perform more types of analyses, such as modeling the effects of clutter or predicting response time. In addition, getting a larger sample of screenshots will allow us to perform statistical analyses on the image processing results as well. Future work will also look to diversify the tasks that physicians perform with the interface.

CHAPTER V

CONCLUSION

This research has provided a comprehensive assessment of the different metrics available to evaluate EMRs. Image processing, GOMS, subjective ratings, and eye tracking were evaluated, and this research provided an overview of the benefits and limitations of the different approaches. In general, certain eye tracking metrics stand out from all other measures in that they correlate best and best reflect the differences in data load between different displays. From a human factors perspective, this is highly-valuable information that could help designers evaluate their displays. GOMS does not appear to be promising, whereas image processing techniques do not reflect human performance, although they can point out usability issues. Subjective information also help point out usability issues but do not reflect the performance effects of data load.

CHAPTER VI

BIBLIOGRAPHY

Ahmed, A., Chandra, S., Herasevich, V., Gajic, O., & Pickering, B. W. (2011). The effect of two different electronic health record user interfaces on intensive care provider task load, errors of cognition, and performance*. *Critical Care Medicine*, *39*(7), 1626–1634. http://doi.org/10.1097/CCM.0b013e31821858a0

Alexander, A. L., Stelzer, E. M., Kim, S.-H., & Kaber, D. B. (2008). Bottom-up and Top-down Contributors to Pilot Perceptions of Display Clutter in Advanced Flight Deck Technologies. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*(18), 1180–1184. http://doi.org/10.1177/154193120805201806

Andre, A. D., & Wickens, C. D. (1995). When Users Want What's not Best for Them. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *3*(4), 10–14. http://doi.org/10.1177/106480469500300403

Asher, M. F., Tolhurst, D. J., Troscianko, T., & Gilchrist, I. D. (2013). Regional effects of clutter on human target detection performance. *Journal of Vision*, *13*(5), 1–15. http://doi.org/10.1167/13.5.25

Baldassi, S., Megna, N., & Burr, D. C. (2006). Visual Clutter Causes High-Magnitude Errors. *PLoS Biology*, *4*(3), e56. http://doi.org/10.1371/journal.pbio.0040056

Beck, M. R., Lohrenz, M. C., & Trafton, J. G. (2010). Measuring search efficiency in complex visual search tasks: global and local clutter. *Journal of Experimental Psychology. Applied*, *16*(3), 238–250. http://doi.org/10.1037/a0019633

Berg, R. Van Den, Cornelissen, F. W., & Roerdink, J. B. T. M. (2009). A crowding model of visual clutter, *9*(2009), 1–11. http://doi.org/10.1167/9.4.24.Introduction

Berry, A. B. L., Butler, K. A., Harrington, C., Braxton, M. O., Walker, A. J., Pete, N., … Haselkorn, M. (2016). Using conceptual work products of health care to design health {IT}. *Journal of Biomedical Informatics*, *59*, 15–30. http://doi.org/http://dx.doi.org/10.1016/j.jbi.2015.10.014

Bobillo, F., Delgado, M., & G??mez-Romero, J. (2008). Representation of context-dependant knowledge in ontologies: A model and an application. *Expert Systems with Applications*, *35*(4), 1899–1908. http://doi.org/10.1016/j.eswa.2007.08.090

Bravo, M. J., & Farid, H. (2006). Object recognition in dense clutter. *Perception & Psychophysics*, *68*(6), 911–918. http://doi.org/10.3758/BF03193354

Card, S. K., Moran, T. P., & Newell, A. (1980). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, *23*(7), 396–410. http://doi.org/http://doi.acm.org/10.1145/358886.358895

Chang, H., Zhang, J., Liu, X., Yang, C., & Li, Q. (2010). Color Image Clutter Metrics For Predicting Human Target Acquisition Performance, 1–4.

Clarke, M. A., Belden, J. L., & Kim, M. S. (2015). What Learnability Issues Do Primary Care Physicians Experience When Doing CPOE. *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*,

9171, 373–383. http://doi.org/10.1007/978-3-319-21006-3
Cowen, L., Baht, L. J., & Delin, J. (2002). Usability.
Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A Random Glance at the Flight Deck: Pilots' Scanning Strategies and the Real-Time Assessment of Mental Workload. *Journal of Cognitive Engineering and Decision Making*, *1*(3), 271–285. http://doi.org/10.1518/155534307X255627.
Duftschmid, G., Rinner, C., Kohler, M., Huebner-Bloder, G., Saboor, S., & Ammenwerth, E. (2013). The EHR-ARCHE project: Satisfying clinical information needs in a Shared Electronic Health Record System based on IHE XDS and Archetypes. *International Journal of Medical Informatics*, *82*(12), 1195–1207. http://doi.org/10.1016/j.ijmedinf.2013.08.002
Ellis, K. K. E. (2009). Eye tracking metrics for workload estimation in flight deck operations. *ProQuest Dissertations and Theses*, *1467693*, 115. Retrieved from http://ezproxy.net.ucf.edu/login?url=http://search.proquest.com/docview/304901025?accountid=10003\nhttp://sfx.fcla.edu/ucf?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+&+theses&sid=ProQ:ProQuest+Dissertations+&+The
Ewing, G. J., Woodruff, C. J., & Vickers, D. (2006). Effects of "local" clutter on human target detection. *Spatial Vision*, *19*(1), 37–60. http://doi.org/10.1163/156856806775009232
Farley, H. L., Baumlin, K. M., Hamedani, A. G., Cheung, D. S., Edwards, M. R., Fuller, D. C., … Pines, J. M. (2013). Quality and safety implications of emergency department information systems. *Annals of Emergency Medicine*, *62*(4), 399–407. http://doi.org/10.1016/j.annemergmed.2013.05.019
Findaly, J. M. (2004). Eye scanning and visual search. In *The interface of language, vision and action: Eye movements in the visual world* (pp. 135–159). http://doi.org/10.4324/9780203488430
Goldberg, J., & Kotval, X. (1999). Computer Interface Evaluation Using Eye Movements: Methods and Construct. *International Journal of Industrial Ergonomics*, *24*(6), 631–645. http://doi.org/10.1016/S0169-8141(98)00068-7
Grahame, M., Laberge, J., & Scialfa, C. T. (2004). Age Differences in Search of Web Pages: The Effects of Link Size, Link Number, and Clutter. *Human Factors*, *46*(3), 385–98. http://doi.org/10.1518/hfes.46.3.385.50404
Hammond, K. W., Efthimiadis, E. N., & Laundry, R. J. (2011). Efficient de-identification of electronic patient records for user cognitive testing. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2771–2778. http://doi.org/10.1109/HICSS.2012.236
Hegarty, M., De Leeuw, K., & Bonura, B. (n.d.). What do spatial ability tests really measure. In *2008, November*. Chicago, IL: 49th meeting of the Psychonomic Society.
Henderson, J. M., & Smith, T. J. (2009). The in fl uence of clutter on real-world scene search : Evidence from search ef fi ciency and eye movements, *9*(2009), 1–8. http://doi.org/10.1167/9.1.32.Introduction
Holy, L., Jezek, K., Snajberk, J., & Brada, P. (2012). Lowering visual clutter in large component diagrams. *Proceedings of the International Conference on Information Visualisation*, 36–41. http://doi.org/10.1109/IV.2012.17
James, J. T. (2013). A new, evidence-based estimate of patient harms associated with hospital care. *Journal of Patient Safety*, *9*(3), 122–8. http://doi.org/10.1097/PTS.0b013e3182948a69
Jansen, M., & Van Kreveld, M. (1998). Evaluating the consistency of cartographic

generalization. *Proc. 8th Int. Symp. on Spatial Data Handling*, (21957), 668–678. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.54.3484

John, B. E., & Kieras, D. E. (1996). Using GOMS for user interface design and evaluation: Which technique? *ACM Transactions on Computer-Human Interaction (TOCHI)*, *3*(4), 287–319. http://doi.org/10.1145/235833.236050

Just, M. A., & Carpenter, P. A. (1978). Inference processes during reading: Reflections from eye fixations. *Eye Movements and the Higher Psychological Functions.*, Hillsdale, N.J.: Erlbaum.157–74.

Karsh, B. T., Weinger, M. B., Abbott, P. A., & Wears, R. L. (2010). Health information technology: fallacies and sober realities. *J Am Med Inform Assoc*, *17*(6), 617–623. http://doi.org/10.1136/jamia.2010.005637

Kim, M. S., Shapiro, J. S., Genes, N., Aguilar, M. V., Mohrer, D., Baumlin, K., & Belden, J. L. (2012). A Pilot Study on Usability Analysis of Emergency Department Information System by Nurses. *Applied Clinical Informatics*, *3*(1), 135–153. http://doi.org/10.4338/ACI-2011-11-RA-0065

Kim, S., & Kaber, D. B. (2009). Assessing the Effects of Conformal Terrain Features in Advanced Head- Up Displays on Pilot Performance. *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting*, 36–40.

Kohm, L., Corrigan, J., & Donaldson, M. (2000). To err is Human: Building a Safer Health Care System. *Institute of Medicine*, 26–48. http://doi.org/10.1017/S0140525X00015685

Lanham, H. J., Sittig, D. F., Leykum, L. K., Parchman, M. L., Pugh, J. a, & McDaniel, R. R. (2014). Understanding differences in electronic health record (EHR) use: linking individual physicians' perceptions of uncertainty and EHR use patterns in ambulatory care. *Journal of the American Medical Informatics Association : JAMIA*, *21*(1), 73–81. http://doi.org/10.1136/amiajnl-2012-001377

Lohrenz, M. C., Trafton, J. G., Beck, R. M., & Gendron, M. L. (2009). A model of clutter for complex, multivariate geospatial displays. *Human Factors*, *51*(1), 90–101. http://doi.org/10.1177/0018720809333518

Miniukovich, A., & De Angeli, A. (2014). Quantification of interface visual complexity. *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, 153–160. http://doi.org/10.1145/2598153.2598173

Moacdieh, N. M., & Sarter, N. B. (2012). Eye Tracking Metrics: A Toolbox for Assessing the Effects of Clutter on Attention Allocation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 1366–1370. http://doi.org/10.1177/1071181312561391

Moacdieh, N., & Sarter, N. (2014). Display Clutter: A Review of Definitions and Measurement Techniques. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 0018720814541145–. http://doi.org/10.1177/0018720814541145

Moacdieh, N., & Sarter, N. (2015). Clutter in Electronic Medical Records: Examining Its Performance and Attentional Costs Using Eye Tracking. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(4), 591–606. http://doi.org/10.1177/0018720814564594

Munn, S. M., Stefano, L., & Pelz, J. B. (2008). Fixation-identification in dynamic scenes. *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization - APGV '08*, *1*(212), 9. http://doi.org/10.1145/1394281.1394287

Murphy, D. R., Reis, B., Sittig, D. F., & Singh, H. (2012). Notifications Received by Primary

Care Practitioners in Electronic Health Records: A Taxonomy and Time Analysis. *The American Journal of Medicine*, *125*(2), 209.e1–209.e7. http://doi.org/10.1016/j.amjmed.2011.07.029

Mušicki, D., Suvorova, S., Morelande, M., & Moran, B. (2005). Clutter map and target tracking. *2005 7th International Conference on Information Fusion, FUSION*, *1*, 69–76. http://doi.org/10.1109/ICIF.2005.1591838

Pereira, E. J., & Castelhano, M. S. (2014). Peripheral Guidance in Scenes: The Interaction of Scene Context and Object Content. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), Advance online publication. http://doi.org/10.1037/a0037524

Poole, A., & Ball, L. J. (2006). Eye Tracking in HCI and Usability Research. In *Encyclopedia of Human Computer Interaction* (pp. 211–219). http://doi.org/10.4018/978-1-59140-562-7.ch034

Rosenholtz, R., Li, Y., Jin, Z., & Mansfield, J. (2005). Feature congestion: A measure of visual clutter. *Journal of Vision*, *6*(6), 827–827. http://doi.org/10.1167/6.6.827

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*(2), 17.1–22. http://doi.org/10.1167/7.2.17

Saitwal, H., Feng, X., Walji, M., Patel, V., & Zhang, J. (2010). Assessing performance of an Electronic Health Record (EHR) using Cognitive Task Analysis. *International Journal of Medical Informatics*, *79*(7), 501–506. http://doi.org/10.1016/j.ijmedinf.2010.04.001

Schraagen, J. M., Chipman, S. F., & Shalin, V. L. (2000). Cognitive Task Analysis, 546. Retrieved from http://books.google.dk/books?id=gyt5AgAAQBAJ

Segall, N., Kaber, D. B., Taekman, J. M., & Wright, M. C. (2013). A Cognitive Modeling Approach to Decision Support Tool Design for Anesthesia Provider Crisis Management. *International Journal of Human-Computer Interaction*, *29*(2), 55–66. http://doi.org/10.1080/10447318.2012.681220

Sills, M. (2015). Adaptive User Interface Based on Eye Tracking.

Singh, H., Spitzmueller, C., Petersen, N. J., Sawhney, M. K., & Sittig, D. F. (2013). Information Overload and Missed Test results in Electronic Health record-based Settings. *Jama*, *167*(2), 187–188. http://doi.org/10.1001/2013

Tasa, U. B., Ozcan, O., Yantac, A. E., & Unluer, A. (2008). A case study on better iconographic design in electronic medical records' user interface. *Informatics for Health & Social Care*, *33*(2), 125–38. http://doi.org/10.1080/17538150802127298

Terry, A. L., Brown, J. B., Bestard Denomme, L., Thind, A., & Stewart, M. (2012). Perspectives on electronic medical record implementation after two years of use in primary health care practice. *Journal of the American Board of Family Medicine : JABFM*, *25*(4), 522–7. http://doi.org/10.3122/jabfm.2012.04.110089

Tully, M. P., Kettis, Å., Höglund, A. T., Mörlin, C., Schwan, Å., & Ljungberg, C. (2013). Transfer of data or re-creation of knowledge – Experiences of a shared electronic patient medical records system. *Research in Social and Administrative Pharmacy*, *9*(6), 965–974. http://doi.org/10.1016/j.sapharm.2013.02.004

Van Vleck, T. T., Stein, D. M., Stetson, P. D., & Johnson, S. B. (2007). Assessing data relevance for automated generation of a clinical summary. *AMIA ... Annual Symposium Proceedings / AMIA Symposium. AMIA Symposium*, 761–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18693939

Weir, C. R., Nebeker, J. J. R., Hicken, B. L., Campo, R., Drews, F., & LeBar, B. (2007). A

Cognitive Task Analysis of Information Management Strategies in a Computerized Provider Order Entry Environment. *Journal of the American Medical Informatics Association*, *14*(1), 65–75. http://doi.org/10.1197/jamia.M2231

Westerbeek, H., & Maes, A. (2004). Referential scope and visual clutter in navigation tasks. *Proceedings of PRE-CogSci 2011*.

Wickens, C. D., & Schons, V. (1993). Visual Separation and Information Access in Aircraft Display Layout.

Wu, L., Zhu, Z., Cao, H., & Li, B. (2015). Influence of information overload on operator's user experience of human–machine interface in LED manufacturing systems. *Cognition, Technology & Work*. http://doi.org/10.1007/s10111-015-0352-0

Yang, K., Member, S., Du, E. Y., Member, S., Delp, E. J., Jiang, P., & Chen, Y. (2013). A New Approach of Visual Clutter Analysis for Pedestrian Detection. *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, *1*(Itsc), 1173–1178.

Yeh, M., Merlo, J. L., Wickens, C. D., & Brandenburg, D. L. (2003). Head up versus head down: the costs of imprecision, unreliability, and visual clutter on cue effectiveness for display signaling. *Human Factors*, *45*(3), 390–407. http://doi.org/10.1518/hfes.45.3.390.27249

Zakaria, S., & Ghani, M. K. A. (2013). The Impact of EMR User Interface Design on Doctor Satisfaction. *E-Proceeding of Software Engineering Postgraduates Workshop (SEPoW)*, 94.

Zelinsky, G. J. (2008). A Theory of Eye Movements During Target Acquisition. *Psychological Review*, *115*(4), 787–835. http://doi.org/10.1037/a0013118

Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel-serial visual search. *Journal of Experimental Psychology. Human Perception and Performance*, *23*(1), 244–262. http://doi.org/10.1037/0096-1523.23.1.244

Zeng, Q., Cimino, J. J., & Zou, K. H. (2002). Providing Concept-oriented Views for Clinical Data Using a Knowledge-based System: An Evaluation. *Journal of the American Medical Informatics Association*, *9*(3), 294–305. http://doi.org/10.1197/jamia.M1008

Zhang, R., Pakhomov, S., McInnes, B. T., & Melton, G. B. (2011). Evaluating measures of redundancy in clinical texts. *AMIA ... Annual Symposium Proceedings / AMIA Symposium. AMIA Symposium*, *2011*, 1612–20. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3243221&tool=pmcentrez&rendertype=abstract

APPENDIX

I. COGNITIVE TASK ANALYSIS

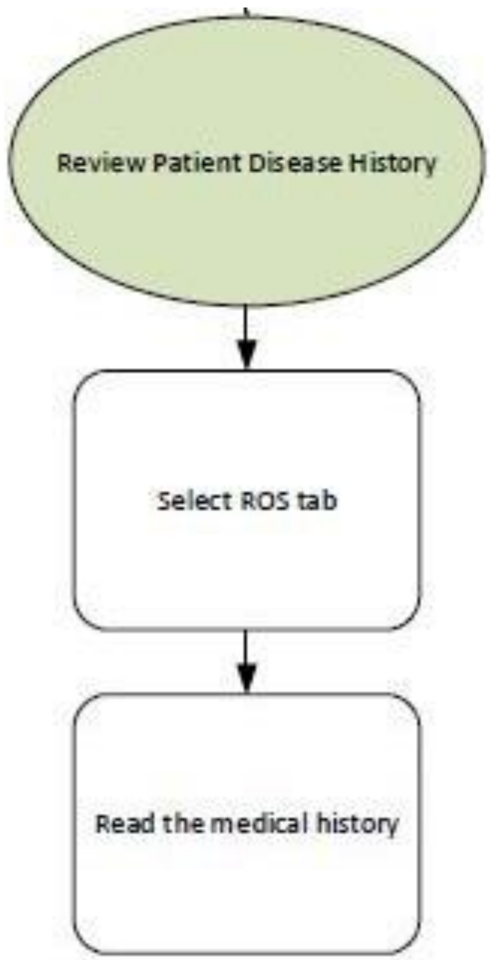
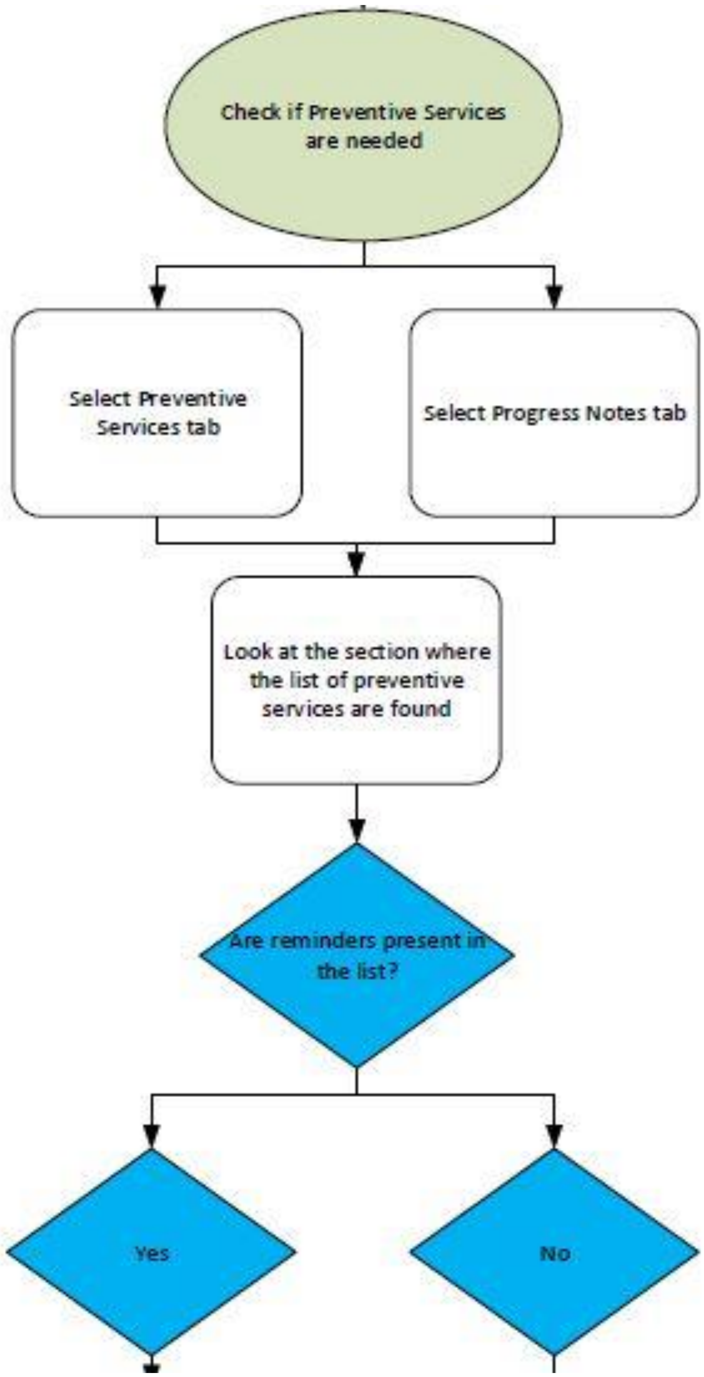
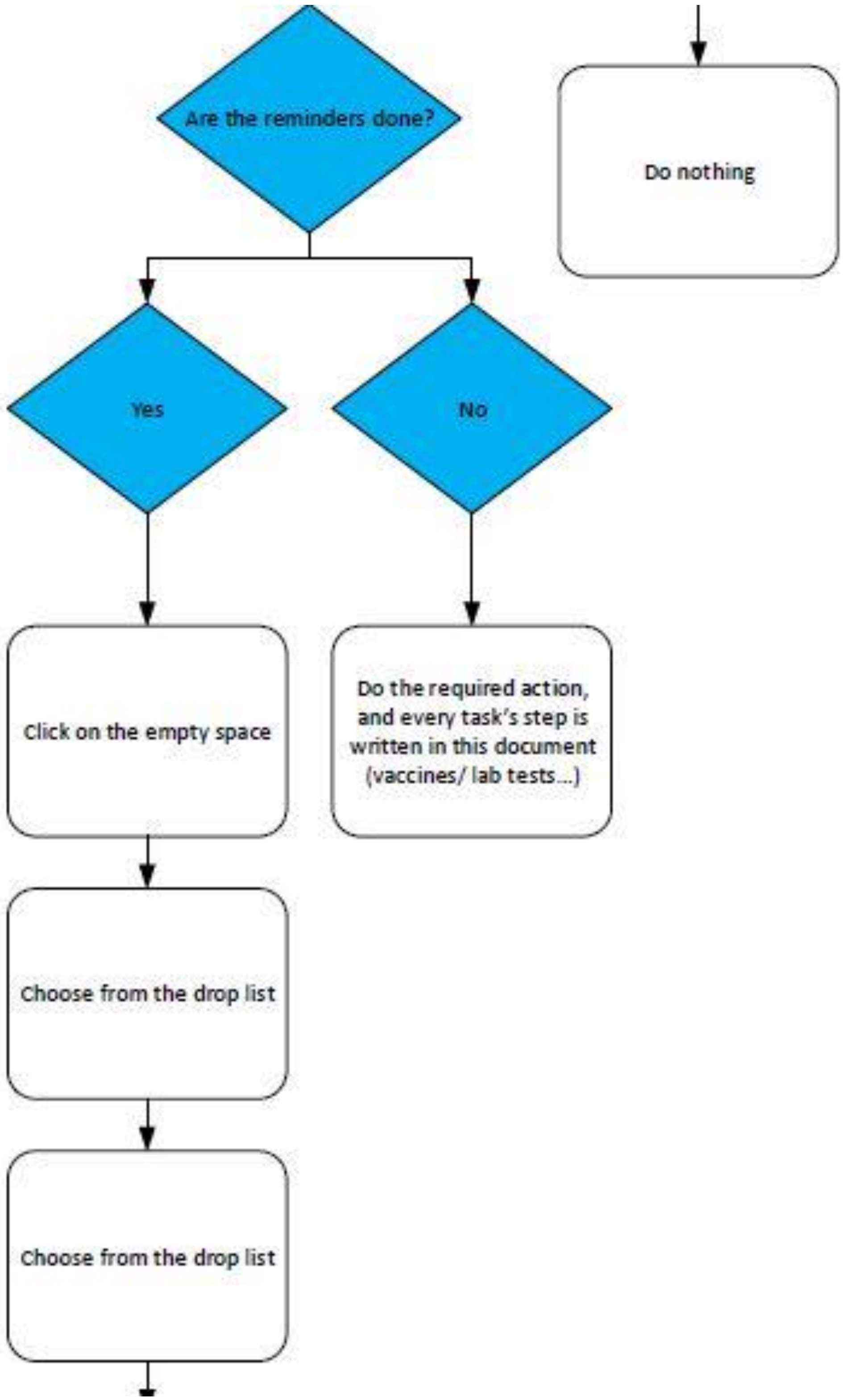


Figure 6. Cognitive Task Analysis for the Second Task



Check if Preventive Services are needed
Select Preventive Services tab
Select Progress Notes tab
Look at the section where the list of preventive services are found
Are reminders present in the list?
Yes
No


Are the reminders done?
Do nothing
Yes
No
Click on the empty space
Do the required action, and every task's step is written in this document (vaccines/ lab tests...)
Choose from the drop list
Choose from the drop list

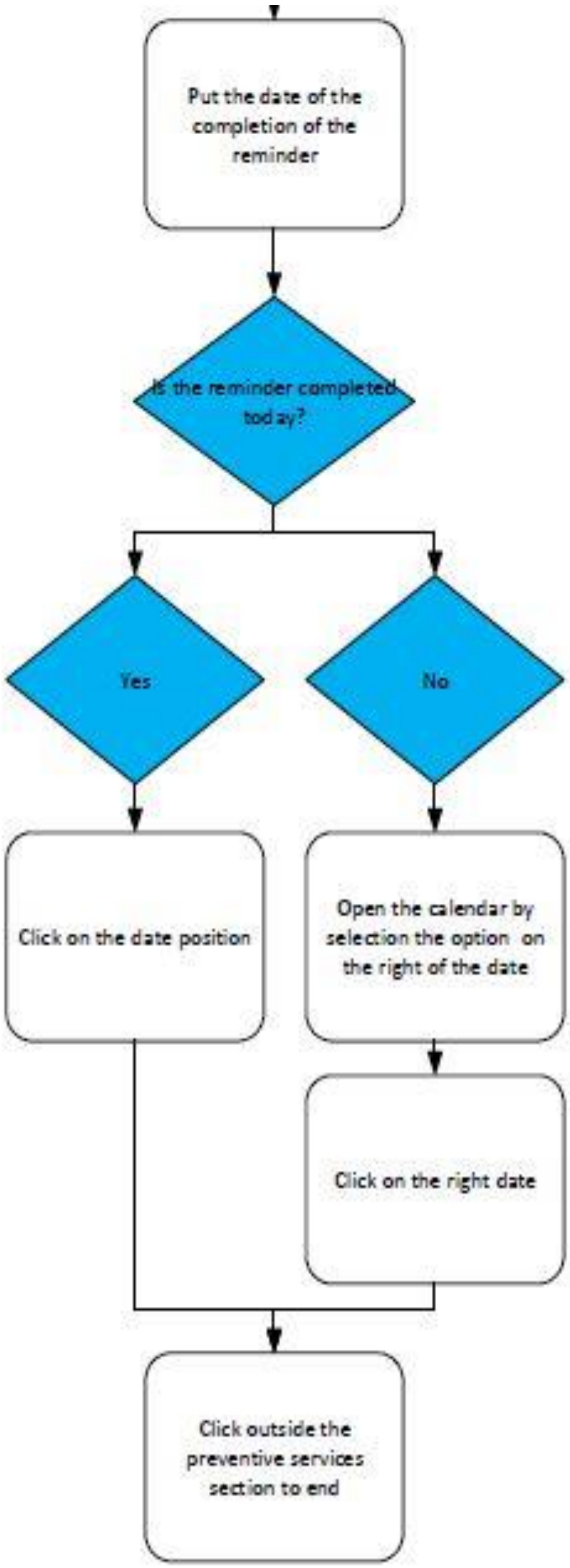


Figure 7. Cognitive Task Analysis for the Third Task

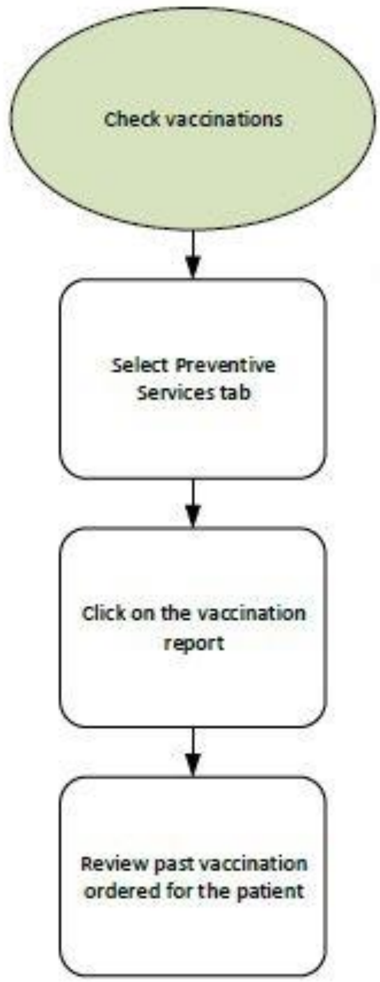


Figure 8. Cognitive Task Analysis for the fourth Task

II. EMR DISPLAYS

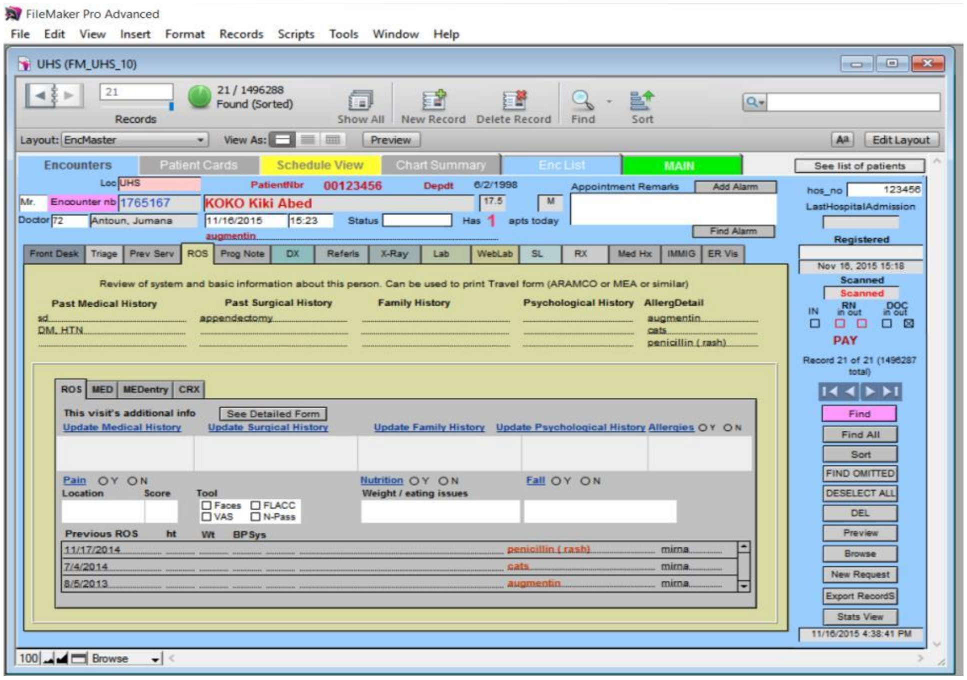


Figure 9. EMR Interface for the ROS Display

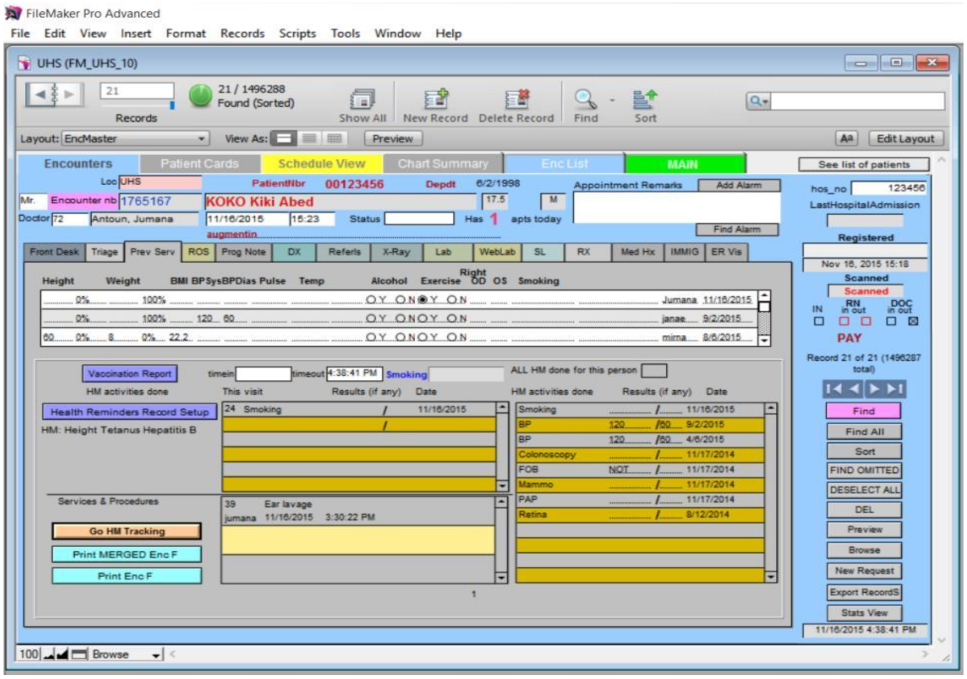


Figure 10. EMR Interface for the Preventive Services Display

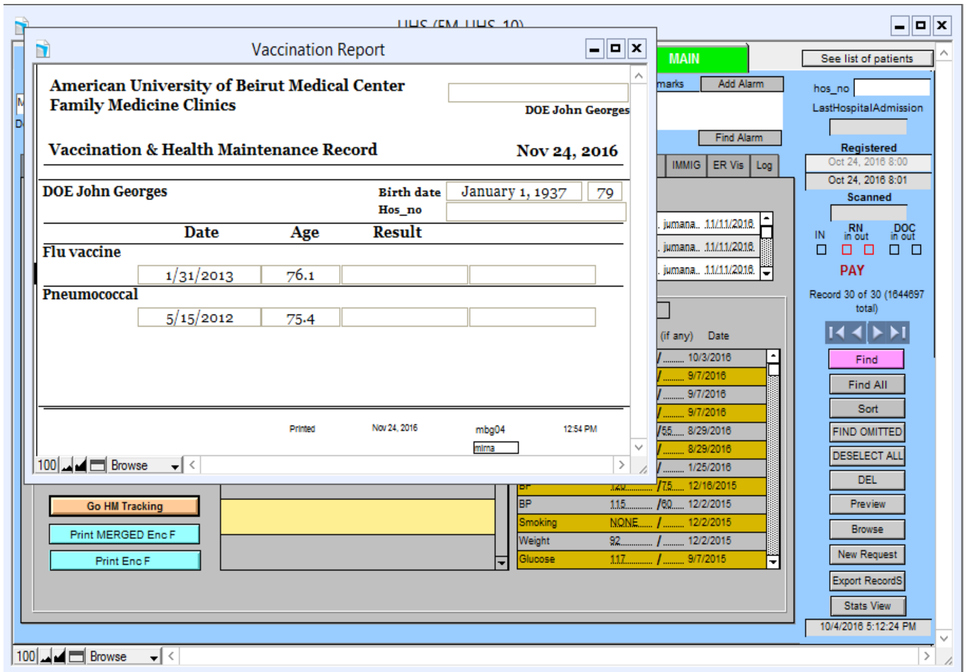


Figure 11. EMR Interface for the Vaccination Display

III. PHASE II FORM

Participant Number & Age: . . . .   Date: . . . .

Participant Experience with the EMR (# of years): . . . .

Display/Task	**Patient #**	**Sequence to complete a task**

Where do you think data overload is present: . . . .

. . . .

. . . .

. . . .

. . . .

IV. PHASE III TASKS

Table 26. Tasks of trial 2

	Tasks for Trial 2: John Doe
1	Check if the patient has done any surgeries. If yes, list the surgeries.
2	You want to check what his last visit was about in August 2016. What was his complaint when he presented in August 2016?
3	The patient is complaining of abdominal pain and weight loss. Check his weight over the past 1 year. What was his weight in Dec 2015?
4	What was the age of the patient when he took the pneumococcal vaccine?

Table 27. Tasks of trial 3

	Tasks for Trial 3: John Doe
1	You want to know if this patient has chronic disease to guide your management plan? Does this patient have Diabetes?
2	The patient is presenting with cough and fever. He reports that he had similar episode in Dec 2015 and he was given a medication that worked a lot. What was the antibiotics he was given?
3	As part of comprehensive care of the patient, you want to know if his blood pressure is controlled below 140/90. Check his previous BP. Is he controlled for the past year?
4	You tell the patient that he needs to take flu vaccine vaccine when he is cured from this infection. He tells you he has already taken the vaccine. Check in his chart if he has taken the vaccine. Did he take the flu shot this year?

Table 28. Tasks of trial 4

	Tasks for Trial 4: John Doe 1
1	Before you ask the patient to come into the room, you want to find his chronic diseases. List his problems.
2	Your patient is presenting for knee pain; he reports that he had similar pain 2-3 months ago, and was given a medication that really helped him. Please find the medication he was prescribed.
3	The patient is worried about recent high blood pressure. Check if his previous readings in the clinic in the last 2 years were high too. Give examples of previous BP readings and their dates.
4	You found out that he works with machines and has risk to cut himself. You want to check if he has taken tetanus vaccine in the past 10 years. Did he take the vaccine?

Table 29. Tasks of trial 5

	Tasks for Trial 5: John Doe 2
1	Before you ask the patient to come into the room, you want to find his chronic diseases. List his problems.
2	The patient tells you that he once complained of hearing loss and you asked for a test. Find that note.
3	Mother is worried that he is not getting taller. Compare his height over the years.
4	Check if he took tetanus at 10 years

Table 30. Tasks of trial 6

	Tasks for Trial 6: John Doe 3
1	Before you ask the patient to come into the room, you want to find his chronic diseases. List his problems.
2	The patient wants to know when she came to the clinic and had influenza. It was after her trip to India.

3	The patient is worried about weight loss. Check her weight over the years. What was her weight in 2013?
4	She wants to know if she has taken Hepatitis A vaccine

V. PHASE III DEBRIEFING FORM

Evaluation of methods for assessing the usability of electronic medical records
Debriefing form (Phase 2)

AUB
Department of Industrial Engineering and Management
Nadine Marie Moacdieh (PI), Jumana Antoun (Co-I), Maher Ghalayini (Co-I)

Participant information

Age: ___________
Years with the AUB Department of Family Medicine: _________
Please indicate if you are a resident: Yes/No

Ratings

How proficient would you consider yourself to be with the current system that you use?
(Not proficient) 1 2 3 4 5 (Very proficient)

How would you rate the amount of data in the current system that you use?
(Very low data load) 1 2 3 4 5 (Very high data load)

How would you rate the amount of mental load involved in using the current system?
(Very low mental load) 1 2 3 4 5 (Very high mental load)

How would you rate the amount of data in the other system that you used today?
(Very low data load) 1 2 3 4 5 (Very high data load)

How would you rate the amount of mental load involved in using this other system?
(Very low mental load) 1 2 3 4 5 (Very high mental load)

In general, which interface do you think is easier to use?

- Current system
- Other, new system

(Optional) If you could adjust some things in the current system that you use, what would it be?
____
____
____

Performance assessment

1. In general, you believe that your performance on your given tasks was
   (very poor)1   2   3   4   5 (excellent)
   (Optional) Please explain any issues that you faced:
   ____

2. In general, the tasks, scenarios, and displays used in this experiment adequately represented real tasks
   1) Yes
   2) No. If no, please explain:
      ____

VI. GOMS RESULTS

Table 31. CTA GOMS of interface 1 task 2 trial 1

Low-Data Display Task 2 Trial 1		
Search for the "Progress Notes" word	M	1.35
Hover hand to mouse	H	0.4
Drag mouse to "Progress Notes" words	P	1.1
Click using the left mouse button	K	0.2
Search for the first note	M	1.35
Drag mouse to the first note	P	1.1
Click using the left mouse button	K	0.2
Check if it is the correct note	M	1.35
Mouse Scroll	K	0.2
Mouse Scroll	K	0.2
Mouse Scroll	K	0.2
Mouse Scroll	K	0.2
Mouse Scroll	K	0.2
Mouse Scroll	K	0.2
Scan the notes with your eyes	M	1.35
	Total	9.6

Table 32. CTA GOMS of interface 1 task 3 trial 1

Low-Data Display Task 3 Trial 1		
Search for the "Preventive Services" words	M	1.35
Hover hand to mouse	H	0.4
drag mouse to "Preventive Services" words	P	1.1

Click using the left mouse button	K	0.2
Search for the date of the services	M	1.35
Drag mouse to the scroll bar	P	1.1
left click on the arrow of the scroll bar	K	0.2
left click on the arrow of the scroll bar	K	0.2
left click on the arrow of the scroll bar	K	0.2
left click on the arrow of the scroll bar	K	0.2
left click on the arrow of the scroll bar	K	0.2
left click on the arrow of the scroll bar	K	0.2
Search for the Blood Pressure	M	1.35
	Total	8.05

Table 33. CTA GOMS of interface 1 task 4 trial 1

Low-Data Display Task 4 Trial 1		
Search for the "Preventive Services" words	M	1.35
Hover hand to mouse	H	0.4
drag mouse to "Preventive Services" words	P	1.1
Click using the left mouse button	K	0.2
Search for the "Vaccination Report" words	M	1.35
Drag mouse to "Vaccination Report" words	P	1.1
Left click on the "Vaccination Report" words	K	0.2
Read the data	M	1.35
	Total	7.05

Table 34.CTA GOMS of interface 2 task 1 trial 1

High-Data Display Task 1 Trial 1		

Search for the "Physician" word	M	1.35
Hover hand to mouse	H	0.4
Drag mouse to "Physician" words	P	1.1
Click using the left mouse button	K	0.2
Search for the Medical History	M	1.35
Hover hand to mouse	H	0.4
Drag mouse to the Medical History	P	1.1
Click using the left mouse button	K	0.2
Click using the left mouse button	K	0.2
Read the note	M	1.35
	Total	7.65

Table 35. CTA GOMS of interface 2 task 2 trial 1

High-Data Display Task 2 Trial 1		
Search for the "Physician" word	M	1.35
Hover hand to mouse	H	0.4
Drag mouse to "Physician" word	P	1.1
Click using the left mouse button	K	0.2
Search for the first note	M	1.35
Hover hand to mouse	H	0.4
Drag mouse to the first note	P	1.1
Click using the left mouse button	K	0.2
Check if it is the correct note	M	1.35
Mouse Scroll	K	0.2
Mouse Scroll	K	0.2
Mouse Scroll	K	0.2
Mouse Scroll	K	0.2

Mouse Scroll	K	0.2
Mouse Scroll	K	0.2
Scan the notes with your eyes	M	1.35
	Total	10

Table 36. CTA GOMS of interface 2 task 3 trial 1

High-Data Display Task 3 Trial 1		
Search for the "Physician" word	M	1.35
Hover hand to mouse	H	0.4
Drag mouse to "Physician" words	P	1.1
Click using the left mouse button	K	0.2
Search for the date of the services	M	1.35
Hover hand to mouse	H	0.4
Drag mouse to the scroll bar	P	1.1
left click on the arrow of the scroll bar	K	0.2
left click on the arrow of the scroll bar	K	0.2
left click on the arrow of the scroll bar	K	0.2
left click on the arrow of the scroll bar	K	0.2
left click on the arrow of the scroll bar	K	0.2
left click on the arrow of the scroll bar	K	0.2
Search for the Blood Pressure	M	1.35
	Total	8.45

Table 37. CTA GOMS of interface 2 task 4 trial 1

High-Data Display task 4 Trial 1		
Search for the "Physician" word	M	1.35
Hover hand to mouse	H	0.4
Drag mouse to "Physician" words	P	1.1
Click using the left mouse button	K	0.2
Search for the "Vaccination Report" words	M	1.35
Hover hand to mouse	H	0.4
Drag mouse to "Vaccination Report" words	P	1.1
Left click on the "Vaccination Report" words	K	0.2
Read the data	M	1.35
	Total	7.45

VII. IMAGE PROCESSING RESULTS

Table 38. Image processing of all the tabs in the EMR

	Tab	Feature Congestion	Sub-band Entropy	Edge Density
Low-Data Display	Front Desk	7.1357	3.6771	0.1017
	Lab A	8.5343	3.9914	0.1344
	Lab B	8.2124	3.9039	0.1279
	Lab Additional	7.655	3.6962	0.1091
	DX	8.7454	3.9592	0.1171
	Medication History	6.8114	3.398	0.0949
	Preventive Services	8.6163	3.9285	0.1256
	Progress Notes	7.4764	3.7015	0.1102
	Referrals	6.8468	3.5102	0.0999
	ROS	7.9533	3.8894	0.1135
	Sick Leaves	7.5164	3.6614	0.1044
	Triage	8.4697	3.8709	0.1114
	Vaccination	8.785	3.8193	0.1136
High-Data Display	Physicians	8.3281	3.9783	0.1173