EHR Test Bench

A Prototype Application for

Quantitative Benchmarking

Of

Electronic Health Record Software

By

Paul B Baker

A Capstone Project

Presented to the Department of Medical Informatics & Clinical Epidemiology

And the Oregon Health & Science University School of Medicine

In partial fulfillment of the requirements for the degree of

Masters of Biomedical Informatics

May 2014

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Master's Capstone Project of

Paul B Baker

"EHR Test Bench
A Prototype Application for
Quantitative Benchmarking of Electronic Record Software"

Has been approved

Capstone Advisor
Michael F. Chiang, MD

Table of Contents

List of Illustrations

Acknowledgements

The author would like to acknowledge the ongoing support of Dr. Michael Chiang, mentor and capstone advisor, and wishes to thank him for his valuable help throughout this project. In addition, he would like to thank the clinicians of Silver Spring Emergency Physicians PC, Silver Spring, Maryland, for volunteering to participate in this research. He also acknowledges his troupe of volunteer actors: Richie Manikat, Nahid Sultana, Kate Maloney, and Arthur Pearlstein, whose group efforts made for relatively accurate and certainly amusing clinical scenarios. Finally, he would like to dedicate this work to his wife Joan and sons Brendan, Nick, and Neil, whose patience and support over the past four years has made this educational effort possible.

Abstract

Benchmarking protocols have long been used to measure the performance of both hardware and software products, although rigorous, quantitative comparisons of electronic health record (EHR) solutions are rare in the healthcare informatics marketplace or the research arena. This study uses a prototype software application, *EHR Test Bench,* to benchmark the performance of EHR software systems. The benchmarking program prompts a test subject to view a video of a simulated patient visit and then measures the time taken to create an electronic record of the visit using a commercial EHR charting system. The videos employ volunteer actors to play the roles of patient and doctor, and were scripted to illustrate simple clinical scenarios common in emergency department practice.

The study sought to quantify two key metrics for each of the scenarios: charting efficiency and charting accuracy. The charting efficiency was defined as running time of the video divided by the total time required to watch the video and finish a completed chart, with a theoretical maximum of 100% (user finishing chart when video finishes). Documentation accuracy was measured by scoring each chart against a standard template of history and exam findings predetermined for each case, with a maximum of 100% which indicates complete capture of all data points.

A total of 10 ED clinicians have taken part in the study to date. Baseline characteristics of the study population were surveyed---users averaged 10.7 years of experience with the EHR system being benchmarked and 100% rated themselves as intermediate or advanced users. Adequate inter-observer reliability among testers was noted on calibration cycles (Cronbach's alpha = .73). Each subject watched three case videos during the session. Charting efficiency varied among users and cases: the first case (ankle sprain) was 37.5% (S.D. ± 11.8%), the second case (finger laceration) was 39.3% (S.D. ± 10.0%), and the final case (appendicitis) was 42.4 (S.D. ± 11.9%). Charting accuracy was likewise variable, ranging from 83% (S.D. ± 9%) for the first to 78% (S.D. ± 8%) for the second to 67% (S.D. ± 19%) for the last case. The small sample size at this point did not permit analysis of any correlation between charting speed and accuracy, or the factors such as provider type (physician versus physician assistant) and typing skills in benchmarking parameters. A software evaluation survey of the application completed after the testing protocol was favorable, with over 50% willing to participate in further studies.

This preliminary report suggests that rigorous, quantitative benchmarking of EHR systems can be performed by end users using inexpensive and easily created study materials. Future studies of this type are urgently needed. Web-based testing, using much larger sample sizes and employing multiple different EHR systems, will allow for more statistically powerful and clinically relevant benchmark testing.

Introduction

Charles Friedman proposed the Fundamental Theorem of Informatics in 2009 and the concept remains a critical motivation behind all research and development in medical informatics.[1] In short, the theorem states that the combination of a human and a computer is or should be more powerful and efficient than a human acting alone.

Graphically stated, Friedman's theorem is deceptively simple:

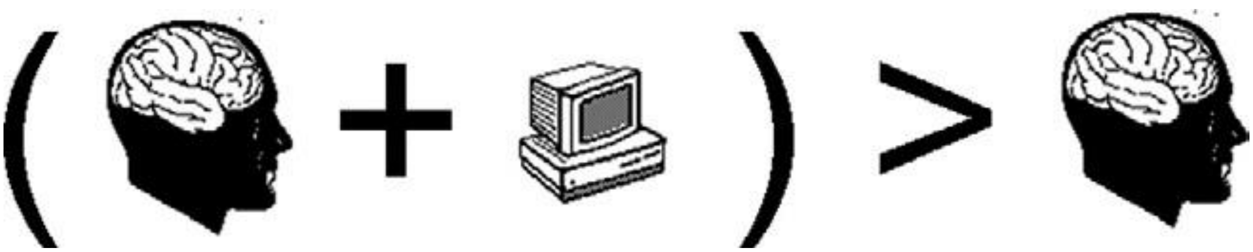


Figure 1 – Fundamental Theorem of Informatics

However if we carry this mathematical analogy further, how do we put numbers into this expression to see if this inequality is indeed true in actual practice? The "human" term in this expression could easily be given one of many numerical values: in today's practice environment, clinicians are being evaluated and rated more than ever before. From patient satisfaction scores to quality measures to productivity figures, the work of individual providers is constantly quantified and compared to their peers and national standards. Regardless of one's opinion on the fairness of these measures, one cannot dispute the utility of "benchmarking" of human resources for the administration of health care services.

On the other hand, substituting an accurate and valid number in the variable for the "computer" term of this statement is more difficult, and in many ways, much more important. The human factor is, mathematically speaking, a constant---the same on both sides of the expression. Gains in efficiency predicted from Friedman's Theorem will be primarily determined by the efficiency of the computer. However, there is a paucity of information when one searches for data on end user "satisfaction scores" or the "productivity figures" for individual electronic health record (EHR) systems used in daily clinical practice. Unlike medical providers, the performance of EHR systems is seldom monitored and evaluated on nationally standardized quality and efficiency measures. For an industry that touts itself as “data-centric”, the lack of hard quantitative performance data of the efficiency and accuracy of EHR systems is both surprising and disappointing. Many frustrated clinicians have often wondered if the inequality in the Fundamental Theorem should be reversed when they struggle to use unwieldy, poorly designed information systems.

Part of the difficulty in evaluating EHR systems is the absence of agreed upon quantitative parameters to benchmark actual performance of these systems in the clinical workplace. In addition, direct measurement of operational efficiency with real patients would interfere with clinical care and raise significant patient privacy concerns. In short, the issue is what exactly do we measure, and what is the best method to make these measurements? The answer to these questions is far from apparent.

What is painfully apparent is the huge sum of money that has been spent on medical informatics software in this country following the Meaningful Use stimulus, with expenditures approaching $40 billion dollars in 2011. [2] Healthcare organizations have made major investments to introduce EHR systems---one study estimated that a 500-bed hospital could spend approximately $50 million over five years to implement an inpatient system.[3] Having made such a significant investment, many hospitals cannot realistically switch to a different system or vendor even if the initial EHR purchased is less than optimal, and therefore have less leverage with an EHR vendor than they would have with other products. Paradoxically, the rush to meet the requirements of Meaningful Use and earn the financial incentives may have led both vendors and hospitals to concentrate on “meaningful use” over true “usability” for clinicians.[4] Kellerman and Jones speculate that “if market forces were allowed to work, doctors might drive vendors to produce more usable products.”[5] A key requirement for such market forces to work is the availability of accurate quantitative benchmarking data on system performance---in order to make an “informed choice” one needs information. Some have advocated for the systematic collection and public reporting of comparative user experiences in order to study and improve the usability and safety of health IT products.[6]

The benefits of the electronic health record (EHR) compared to paper based charting are well-known and agreed-upon: legibility, portability, scalability, clinical decision support, and most importantly, the promise of having vast amounts of clinical data in structured format---the “Holy Grail” of researchers and policymakers alike. Nevertheless, we must balance these benefits against the frequent complaints of end users who believe that EHR systems place too great a load of clerical, “data-entry” tasks on clinicians and take time away from the patient bedside. A recent time study of emergency department physicians calculated that the typical clinician averaged nearly 4000 mouse clicks in a single 10 hour shift.[7] In addition, many physicians complain that documents produced by an EHR templates lack the clarity and readability of traditional handwritten or dictated reports.[8]

In order to improve the usability and performance of EHR systems, we must have a way to quantify the performance of a given EHR using parameters that are most important to clinical end users: speed of document creation and the accuracy of the final chart created by the system. In short, we need a reliable and valid way to benchmark current EHR software. Software benchmarking is defined as "a continuous process whose aim is to improve software products, services, and processes by systematically evaluating and comparing them with those considered to be the best.[9] The "Consumer Reports" of EHR systems has yet to be published, but many authors have identified the need for specific protocols for benchmarking medical information systems.[10]

The first goal of this study will be to develop a prototype software application, the *EHR Test Bench*, which will allow researchers and motivated clinical users to benchmark EHR systems in a standardized, inexpensive, and easily reproducible manner. Because of difficulties inherent in observation of actual patient visits, the study protocol will utilize videotaped patient visits; such simulated patients are widely used for teaching and research purposes in medical education.[11] The *EHR Test Bench* software will present the test subject with a preselected series of videos of simulated emergency department (ED) visits, and prompt the subject to mark the beginning and end times of the charting process. The application will automatically record and tabulate the speed of the charting process for a given video scenario and permit calculation of the speed and efficiency of documentation for a group of users.

The second goal of this research project will be to use the *EHR Test Bench* software to conduct actual benchmark testing of a commercial EHR charting system, using a group of emergency medicine specialists in a large community hospital practice. The benchmarking session will have three phases. Initially, each test subject will be asked to perform a brief series of timed tasks in response to verbal commands delivered in video format, specifically clicking a series of buttons in a particular order. In doing so, we will test the reliability and usability of our testing paradigm and the *EHR Test Bench* application. The test subject will then watch a series of simulated ED visits and create charts using the EHR from the group's practice. We will record the charting time for each scenario and calculate the average charting "speed" for the group for each ED case. Finally, we will analyze the EHR chart created by each user by comparing the data recorded in the final document to a predetermined list of standard data points from each case, and calculate a score for charting "accuracy" for the group as a whole.

The third and final part of the protocol will be a brief software evaluation survey that each test subject will be asked to complete, once finishing the benchmarking protocol. The results of the survey will help us improve the benchmarking application and testing procedures.

We will analyze user preferences in order to refine our testing protocol, and use this feedback to determine the ideal number and length of test cases and benchmarking sessions.

The task of building a fully featured suite of programs for benchmarking EHRs will be a very extensive and long term effort, and will require an iterative process to develop the measurement tools that we so urgently need in informatics research and development. We believe our study is a first step in this direction.

Current Evaluation Methods

Before proceeding further, let us review the methods of software evaluation currently used for EHR systems. The current market for EHRs in the US offers a large number of different software products; The Healthcare Information and Management Systems Society (HIMSS), the major industry group in healthcare informatics, had over 1,100 vendors exhibiting at the 2012 convention.[12] The larger marketplace for hardware and software IT products has long relied upon independent reviews and head-to-head comparisons to permit the customer to make an informed purchase decision. For example, if looking to purchase a video card for your home computer, benchmarking test results (Figure 2) are readily available online:

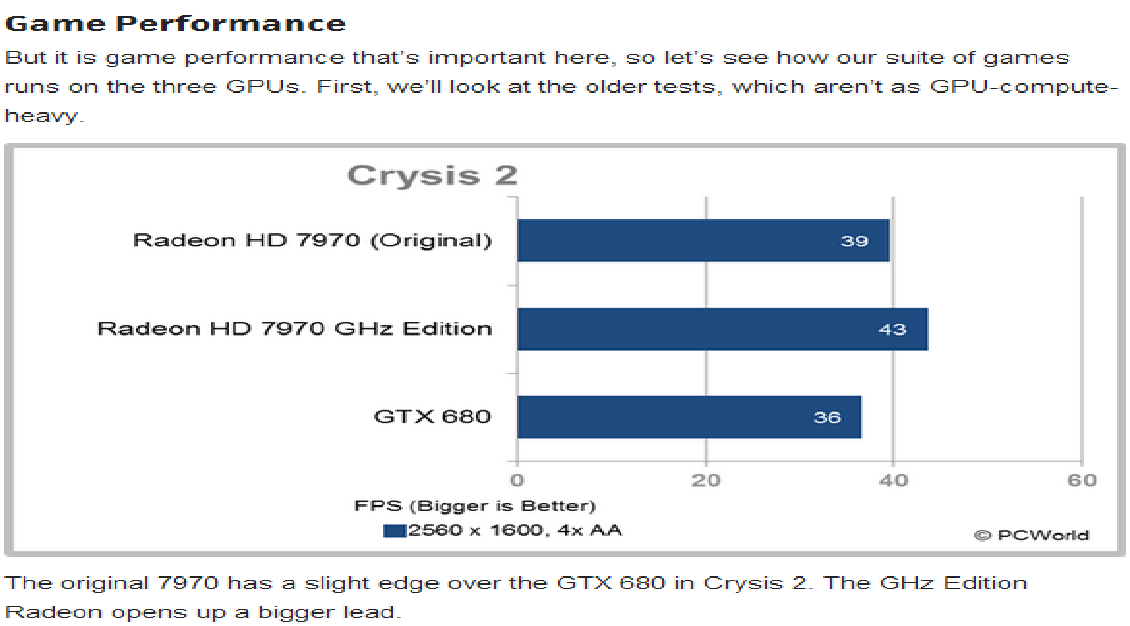


Figure 2 – Results of three competing video cards on video game performance. [13]

While few would equate the results of testing a single hardware component with a thorough evaluation of an enterprise level medical software system, the absence of rigorous, publically available, quantitative data on performance and usability in current EHR systems is both striking and disappointing.

The prospective purchaser or current owner of an EHR system has several options to gain further information about the capabilities of a given software solution. The first source of information about an EHR would be the developer, or in the case of a commercial product, the vendor. While all modern software development teams employ considerable resources in determining user requirements and performance testing both before and after a product release, this data is rarely if ever shared with customers before purchase of the system.

Those seeking information on a particular product could look to vendor-supplied documentation and sales material, which is usually a sales or promotional brochure that describes the product on a very superficial level. A typical example (Figure 3) is the web page description for FirstNet® an emergency department EHR system from Cerner:

Figure 3 – Marketing materials from vendor website[14]

This product description illustrates the point that there is little quantitative information describing actual performance parameters for either clinical or administrative end users. The text focuses instead on stock phrases like "providing safe care efficiently" to promising users can "quickly complete documentation", goals that few would argue against. Even the slickest automobile commercials proudly boast of the latest consumer survey results and list performance data to permit comparisons with other companies' models.

One option for a more balanced evaluation of an EHR system would be independent certifying bodies set up to determine compliance with published requirements for health IT products. The Certification Commission for Healthcare Information Technology (CCHIT), which is a non-profit organization founded in 2004 to "establish the first, comprehensive, practical definition of what capacities were needed [in EHRs and to develop] certification criteria…through a voluntary, consensus-based process engaging diverse stakeholders."[15] CCHIT has created an extensive set of testing procedures that inspect EHR functionality, interoperability and security using criteria developed by expert panels.

Using CCHIT as a model, the Office of National Coordinator for Health Technology has fostered the creation of Authorized Testing and Certification Bodies (ATCBs) that have been approved by the federal government to perform testing of EHRs to certify whether or not a system successfully meets the requirements of the meaningful use (MU) criteria.[16] If a health care provider or organization successfully deploys an ATCB-certified EHR, they are eligible for substantial incentive payments from Medicare.

Nevertheless, it is very important to realize that certifying bodies like CCHIT and other ATCBs only certify that a tested EHR has met the minimal standards to receive MU incentive payments and tells little of the relative usability of one product versus another or its actual performance in the clinical workplace. While qualifying for MU incentive payments can provide additional revenue to a provider or hospital, having a "certified" EHR does not guarantee that such a system will help the individual provider in terms of efficiency and usability.

If certifying agencies do not provide comparative benchmarking data for EHRs, can private consulting groups fill the void? Research firms like Gartner, Inc. are well known in the information technology (IT) industry for the quality and rigor of their research and the resources available to study a problem.[17] Within the healthcare IT field, KLAS Research has provided detailed product evaluations and performance ratings of EHR software to providers and vendors.[18] KLAS conducts extensive product satisfaction surveys of key healthcare IT market segments such as emergency department information systems (EDIS) and radiology imaging software (Picture Archiving and Communications Systems – PACS). Considerable effort is taken to ensure unbiased reporting and detect fraudulent user satisfaction data.

KLAS reports precisely define overall satisfaction with a given EHR or hospital software system; however, several factors limit the usefulness of the KLAS information for product benchmarking. First, KLAS reports represent careful verification and summation of user satisfaction data only, and no actual laboratory or field testing is done by KLAS researchers. Second, KLAS preferentially interviews providers in leadership or executive positions, such as chief information officers (CIO) or chief medical information officers (CMIO).[19] While the views of senior leaders are important, the results may be less significant for the vast majority of clinical end users who perform the bulk of data entry into an EHR. Finally, the information available in KLAS reports is quite expensive, though some data is made available free of charge to health providers who register with the website.

The academic research community has tried to fill this information void and provide reliable data regarding the efficiency and performance of EHR systems in clinical practice. Over the past three decades, numerous studies have been undertaken in an effort to evaluate health IT products. Several large reviews of evaluation studies in the medical informatics literature have been undertaken.[20 21] Yen and Bakken conducted a review in 2012 and identified numerous methodological issues in evaluation studies, including that a "majority of publications…lacked an explicit theoretical framework/model…"[22] As a practical matter, most evaluative studies on EHR systems do not focus on the performance characteristics of specific commercial products and would not be helpful in making purchasing decisions or giving a product development team recommendations for software enhancements.

With regard to research in using EHR systems to document care, most studies to date have consisted of either surveys or time and motion studies. Numerous studies have confirmed that clinicians spent a significant portion of their work day documenting patient care, irrespective of if they are using paper or electronic records. In a survey study by Oxentenko *et.al.*, over two thirds of internal medicine residents reported spending more than four hours per day in documentation while only one third spent this much time in direct patient contact.[23] Another study of hospital based physicians found that as much time was spent on documentation as was spent on direct patient care.[24] A time and motion study in the emergency department showed that thirty to forty percent of physician time is spent in charting and documentation.[25] Similar studies of nurses have estimated that documentation tasks can take up to 35% of their clinical workday.[26]

Given the significant amount of time spent by clinicians in documenting care, has electronic charting been shown to improve the speed or efficiency of documentation? There are not a large number of studies that directly examine the specific effects of EHRs on clinical documentation. A systematic review in 2005 by Poissant *et.al.* looked at the effect on electronic health records on the "time efficiency" of physicians and nurses and found only 23 studies that met inclusion criteria. The review found a decrease in charting time for nurses but an increase in time for physicians, and suggested that "a goal of decreased documentation time in an EHR project is not likely to be realized."[27] One would anticipate that newer studies with more advanced charting systems might show productivity gains. However two recent time and motion studies of the effects of EHRs on nursing documentation time have not demonstrated significant differences between electronic versus traditional paper methods.[28 29] A commercially available ambulatory EHR was shown to increase physician productivity in terms of increased volume of patient visits per month; whether or not this was due to decreased charting time is not clear.[30] This previous study is a rarity in the literature---a quantitative analysis of the effect of an EHR on a group practice using commonly accepted benchmarks such as relative value units (RVUs) and physician productivity measures.

Evaluation of documentation speed is just one component of measuring overall EHR usability. The usability of electronic record systems has become a major area of research in clinical informatics, both to justify the huge investment in health IT and realize the full potential of this technology. Zhang and Walji have pointed out the "lack of EHR-specific usability frameworks and methods" and advocate for a "unified framework of EHR usability."[31] In keeping with the effort to objectively and systematically evaluate products, they have developed software applications to quantify EHR usability such as the Turf EHR Usability Toolkit (Figure 4).

Figure 4 – Usability software application for EHR evaluation.[32]

Other investigators have proposed novel methods to better evaluate the EHR usability. Landman *et.al.* have explored using medical simulation laboratories as a testing environment for evaluating EHR software.[33] Zheng *et.al.* have analyzed navigational patterns in EHR systems by automatically recording the patterns of user interaction with the system (e.g. mouse clicks). By collecting this usage data, patterns of usage can be quantified and workflow efficiency could be optimized.[34] Hripcsak and his colleagues used EHR audit logs to measure the length of time taken to create clinical notes and the time spent in viewing/reading those notes. Interestingly, over 15% of notes created in the EHR at this site were never viewed by anyone.[35]

From this brief review of the literature of EHR evaluation, it is clear that a foundation for the systematic and objective testing of these software systems has been built. However it is equally clear that much work remains to be done.

A New Framework for User-Centered Evaluation

As seen in the previous section, significant work has already been done in developing the methodology to evaluate EHR systems. However with the increasing prevalence and sophistication of healthcare IT products, there is a corresponding need to expand the scope and intensity of the software evaluative process. In short, we must develop a new framework and better techniques to measure the performance of EHR software---and undertake what I define as “quantitative benchmarking” of clinical documentation systems.

In the data-centric world of health information systems, there is much to measure. Any number of parameters of an EHR would be and are excellent candidates for quantitative analysis---system cost, return on investment, and provider productivity, among others. Benchmarks for system response time or frequency of downtime events, for example, are extremely useful to management and technical support personnel. However, I would argue that the principal focus of benchmarking efforts should be on the clinical end users, the stakeholders that are most impacted by EHR performance.

Clinical end users are looking for reliable, user-friendly EHR systems that will facilitate care and not complicate their already busy workday. However, ease of use and facilitation of medical care are hard to quantitate. Most providers would agree that when charting a clinical encounter, two metrics are paramount: speed of document creation and the accuracy of the final document. Fortunately, speed and accuracy are relatively easy to measure and allow comparison to benchmark values.

For these reasons, we have chosen the user-centric values of documentation speed and documentation accuracy as the key metrics of our evaluation framework. Before proceeding, we need to define following key terms:

<u>Clinical information transfer:</u> In this study, we will define this theoretical construct as the flow of information from a dynamic clinical event (e.g. taking a patient’s history) to a static clinical document (e.g. a medical record). This data is generated by the patient-provider interaction, interpreted, summarized, and prioritized by the provider, and ultimately substantiated into a clinical document. This information stream has multiple attributes: speed of transfer, level of detail, data source (patient, family member, review of old records), and accuracy, among many others. This nature of the data flow is highly dependent upon both the patient (by which information he/she chooses or is able to provide) and the clinician (by what he/she chooses to record as significant). The critical factor for informaticians is the modulating

effects of charting aids (pen and paper, dictation, physician scribes, and most importantly, electronic health records) on clinical information transfer.

Encounter time: In this study, the encounter time will be a set of time intervals between the beginning and end of the predefined sub phases of a physician patient encounter. For example, the elapsed time between when a clinician starts and finishes the HPI ($\mathbf{T_{HPI}}$) ---would include the physician's questions and the patient's answers. The total time of the entire encounter ($\mathbf{T_{TOT}}$) is the sum of all the times of the sub phases ($\mathbf{T_{HPI}} + \mathbf{T_{PMH}} + \ldots$).

Documentation time: In this study, the documentation time will be the time of the physician patient encounter ($\mathbf{T_{TOT}}$) plus the additional time the subject uses to complete his/her documentation, the sum of which will be the total documentation time ($\mathbf{DT_{TOT}}$).

Charting efficiency: In this study, charting efficiency ($\boldsymbol{\eta_{Chart}}$) will be the ratio of total encounter time ($\mathbf{T_{TOT}}$) over the total documentation time ($\mathbf{DT_{TOT}}$) expressed as a percentage such that:

$$\boldsymbol{\eta_{Chart}} = \mathbf{T_{TOT} / DT_{TOT} \times 100}$$

By this measure, "perfect" clinical information transfer would result in a charting efficiency of 100%, in which case the total encounter time would be equal to the total documentation time. In this idealized situation, when the provider finished his/her interview of the patient, the documentation would be finished at the same instant. A more typical value, in this investigator's experience, would be where the clinician would spend fifteen minutes with the patient and then an additional ten minutes charting the encounter---with a $\boldsymbol{\eta_{Chart}}$ measuring at **15/(15+10)** or **60%**.

Documentation accuracy: In this study, this documentation accuracy ($\mathbf{A_{CHT}}$) is a percentage measure that compares the data points present in the final charting document with data points present in a gold-standard template created by review of the actual clinical encounter. This gold-standard dataset for a given clinical encounter could be determined prospectively in the case of a simulated encounter or retrospectively in the case of a real-time encounter.

The definition and calculation of a numerical measure for documentation accuracy can be complicated. For readily agreed upon findings, such as those found in the physical examination, it is relatively easy to quantify the presence or absence of a specific exam finding and calculate the completeness of the documentation. The authors of a study comparing paper versus electronic charting of the ophthalmic examination in three common eye disorders developed a scoring algorithm based on the number of findings documented out of the total number possible and reported the percentage as a measure of documentation completeness.[36]

However this approach becomes more problematic when trying to quantify the accuracy of the documentation of the patient history because of the myriad ways that the historical narrative can be interpreted and recorded. One advantage of using scripted and pre-recorded clinical encounter videos is that a transcript of the interview can be analyzed and a dataset of history and physical findings can be collected and agreed upon in advance. The dataset for a given encounter can then be used to create a scoring template for measuring the accuracy of charting documents generated by test subjects.

Hogan and Wagner advocated calculating accuracy by measuring two parameters: correctness (the percentage of recorded findings that are "correct") and completeness (the percentage of recorded findings that are "present").[37] Logan *et.al.* described a method of classifying data elements in a charting document into four categories: correct, incorrect, missing, and extra. These four elements were then used to calculate "correctness" and "completeness" as measure of documentation accuracy.[38] A recent study of the accuracy of ophthalmology documentation utilized two measures of accuracy, "sensitivity" and "positive ratio". The sensitivity was defined as the number of data points documented in the chart divided by the total number of findings present in the gold standard. Positive ratio was defined as the number of true positive findings documented in the chart divided by the number of positive findings in the gold standard.[39]

The scoring algorithm for determining documentation accuracy in this protocol is relatively simple by design. The chart created for each video scenario will be compared to the scoring template for that case (Appendix C - Sample scoring template). For each finding correctly recorded in the chart and present in the template, the user will be awarded a point (+1 – Correct finding). For each finding not recorded in the chart but present in the template, the user will be scored a zero (0 – Absent finding). For each finding incorrectly recorded in the chart but present in the template, the user will be deducted a point (-1 – Incorrect finding). The documentation accuracy score ($\mathbf{A_{CHT}}$) for a chart would then be the percentage score out of the total possible score.

While it is tempting to penalize a charting document for adding additional, possibly correct findings to the chart that were not presented in the scenario video or listed in the scoring template, developing an algorithm to quantify this "extra" data is difficult. Most modern EHR systems allow the user to use pre-populated templates and "copy and paste" techniques to rapidly create "complete" and "billable" documentation, and clinical users rely on such methods to maintain workplace efficiency. Forcing users not to use these methods in our protocol would lessen the validity and real world utility of the benchmarking results obtained. Many authors have commented that using templates and copying techniques have lessened the readability and trustworthiness of EHR records.[40 41] Future studies of the quality and accuracy of EHR documentation will have to account for the presence of superfluous material in the chart and any detrimental effects on clinical communication.

Having outlined a framework for conducting a study of the speed and accuracy of documentation for a given EHR system, what is the most efficient manner to perform the actual benchmarking? A traditional survey study will not provide the requisite quantitative data that we need to make statistically valid comparisons. Direct observation of patient care in a field study would permit some useful measurements but has several major drawbacks. First, benchmarking studies involving real-time patient care can be intrusive and raise significant patient privacy concerns. Second, since every physician-patient encounter is unique and occurs only once in real life, there is no way to standardize the clinical information obtained and directly compare those results to other clinical cases or EHR systems. Finally, studies in live clinical settings are expensive and labor intensive.

We turn instead to the laboratory setting: using simulated patient encounters as a proxy for actual patient visits. Patient simulation has become a valuable component of medical education and is readily adapted to the testing of EHR software.[33] Creating simulated patient encounters has several important advantages: 1.) the exact clinical content present in the interaction can be controlled, 2.) the encounter is standardized and can be used repeatedly for multiple test subjects, and 3.) there are no concerns regarding patient privacy or disruption to patient care and clinical operations.

However, medical simulation facilities are rare outside of academic medical centers and it can be difficult to schedule large numbers of test subjects to visit such facilities. While videotapes of patient encounters lack the realism of live actors or expensive simulator models, such videos are relatively inexpensive, easy to produce, and open up testing of a much greater number of users. Hence, this research protocol will utilize a series of videos of simulated provider-patient encounters illustrating common clinical scenarios. In addition, we will attempt to streamline the benchmarking process by presenting the case scenarios via a software application that will both deliver the videos to the test subject and automatically record the time taken to chart on a given EHR software system. The remainder of this paper will describe the development of prototype software application, *EHR Test Bench*, and its functionality. We will then detail the deployment of this software in a benchmarking study of EHR software used in a community emergency department. We will report preliminary results for documentation speed and accuracy for a series of clinical scenarios commonly seen in emergency medicine.

Description of the *EHR Test Bench* Application

The software program developed for this study is entitled the *EHR Test Bench* as a reflection of the aim of creating a prototype application to automate and quantitate the performance of EHR systems. The principle requirements for the software were:

1.) Allow individuals to logon to the benchmarking session in a secure password-protected manner that maintains the anonymity of the test subject.

2.) Collect survey data from test subjects including years in practice, type of provider, specialty, and other pertinent data such as experience with computers and with the EHR being studied.

3.) Prompt the test subject to view a series of prerecorded videos of patient-provider clinical scenarios selected by the investigator.

4.) Automatically mark the time that a test subject begins a scenario video and finishes his/her charting process in the EHR, and record the total documentation time ($\mathbf{DT_{TOT}}$) in a data file.

5.) Have an intuitive and easy to use user interface (UI) that will not distract the test subject from his/her primary task of watching the case video and charting with the EHR system being studied.

6.) Be written in a coding environment that would allow a novice programmer to rapidly develop a working prototype and easily modify the application without extensive rewriting of code.

Visual Studio 2012, an integrated development environment (IDE) from Microsoft was chosen as the platform for development of the application. Code was written in Visual Basic and the runtime was installed on a Dell XPS 18 All-In-One touchscreen computer running Windows 8.1.

Although the application could have been designed to simultaneously test a group of test subjects in a classroom setting or over a network, limitations in programming expertise forced us to design the application to function in a single user format. In this design paradigm, the investigator setups the testing station in the *EHR Test Bench* in advance for each individual test subject using a predetermined set of clinical scenario videos from our benchmarking protocol.

The screen sequence the study investigator and test subject would see in a walkthrough of the application is fairly straightforward. The test proctor would initially sign-in to the program using an administrative user ID and password using the following screen (Figure 5).

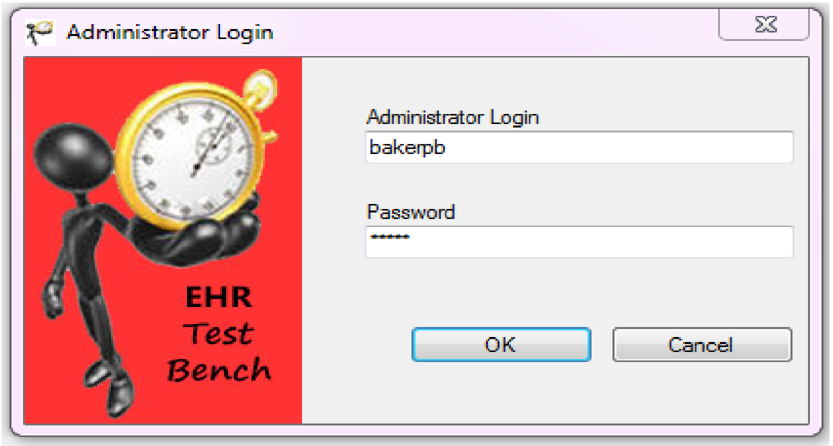


Figure 5 – Administrative login screen.

After successfully logging in, the test proctor enters the details of the protocol (Figure 6).

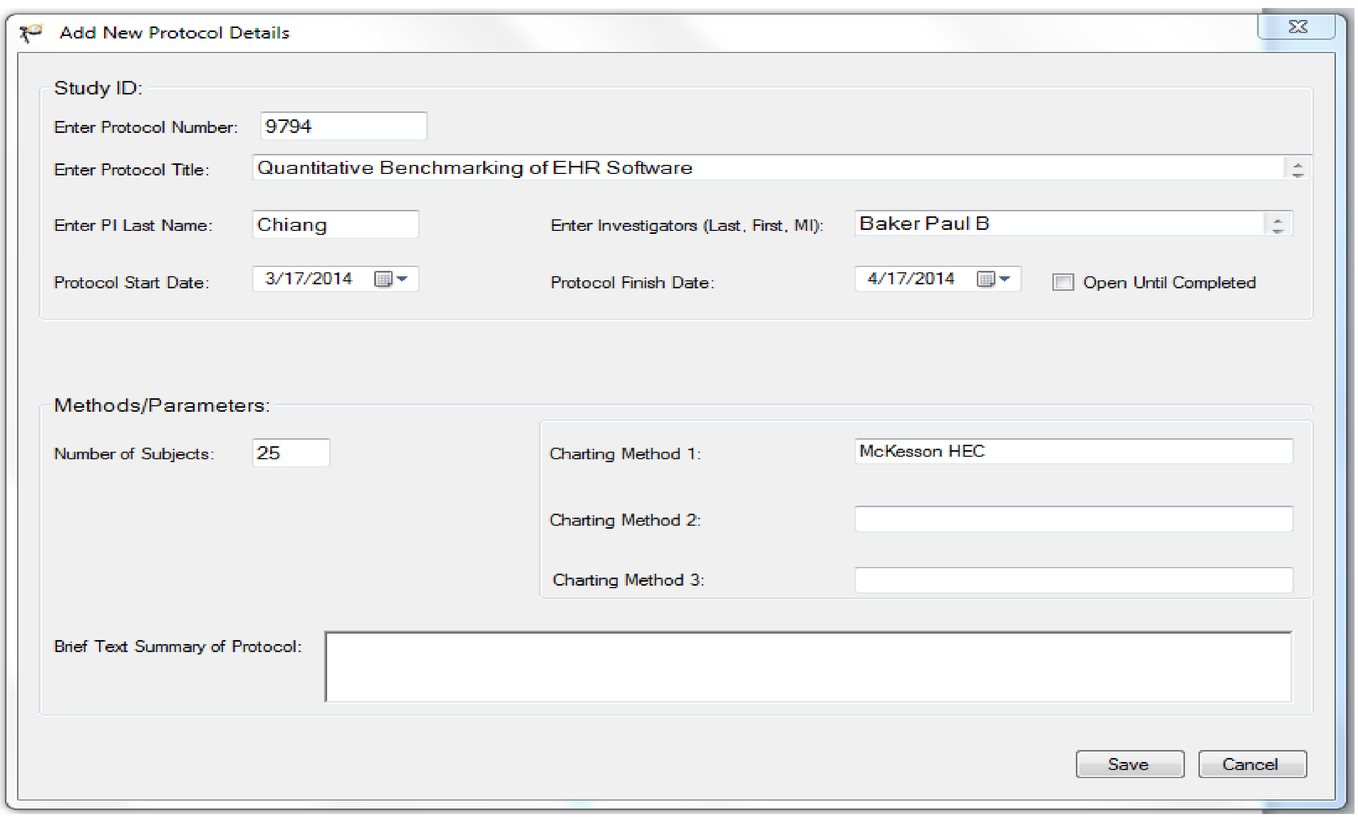


Figure 6 – Add New Protocol Details screen.

Protocol details are reflected in the Administrative Test Setup screen (Figure 7).

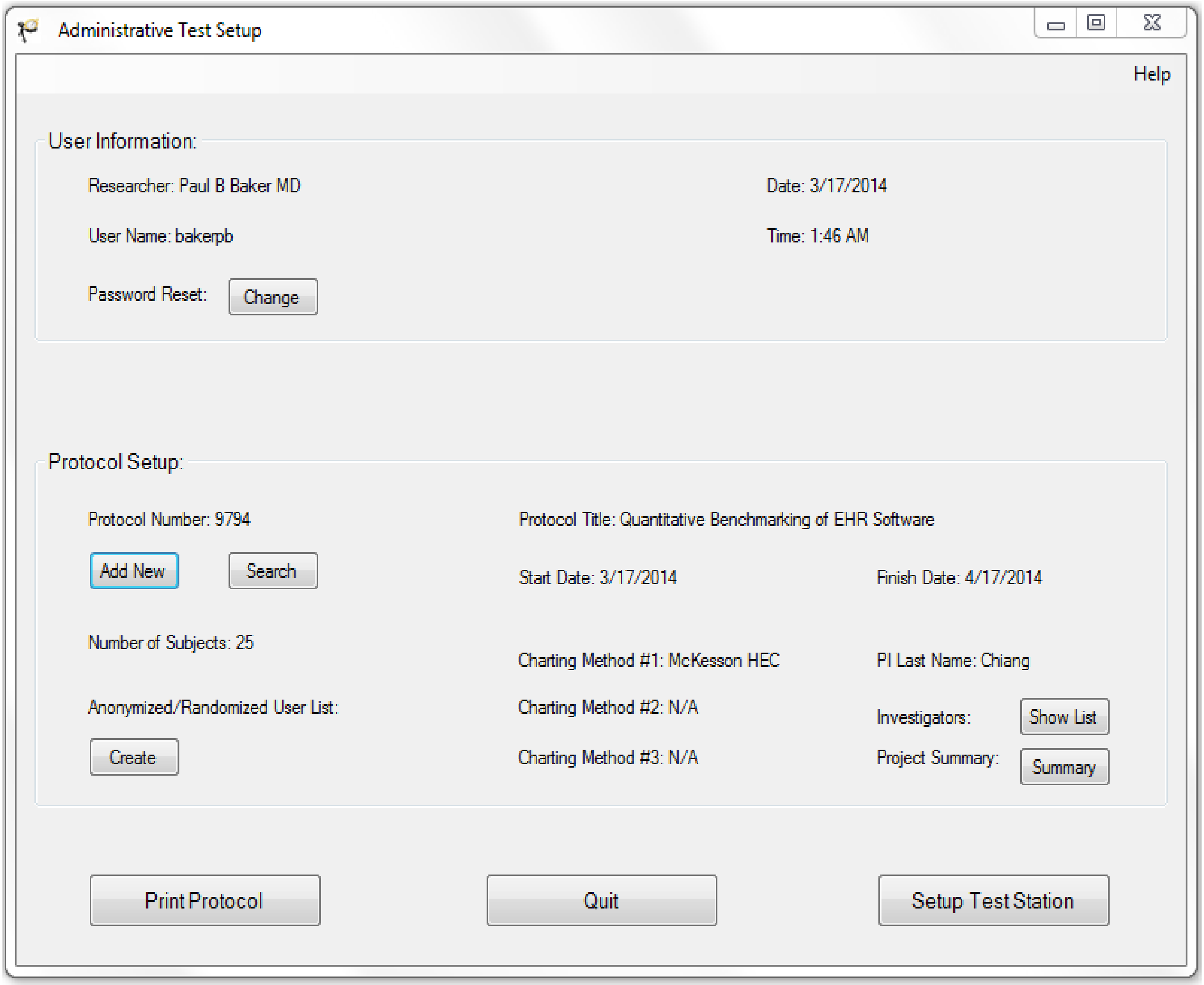


Figure 7 – Administrative Test Setup screen.

The application has the capability to allow comparative benchmarking of up to three different methods of charting simultaneously. In this early phase protocol, only one charting method (an Emergency Department EHR from McKesson) will be studied due to a relatively small number of test subjects available.

Another feature of the software is the capability to generate a list of randomized numbers and print out a series of user ID and password combinations to allow anonymization of the identity of test subjects (Figure 8).

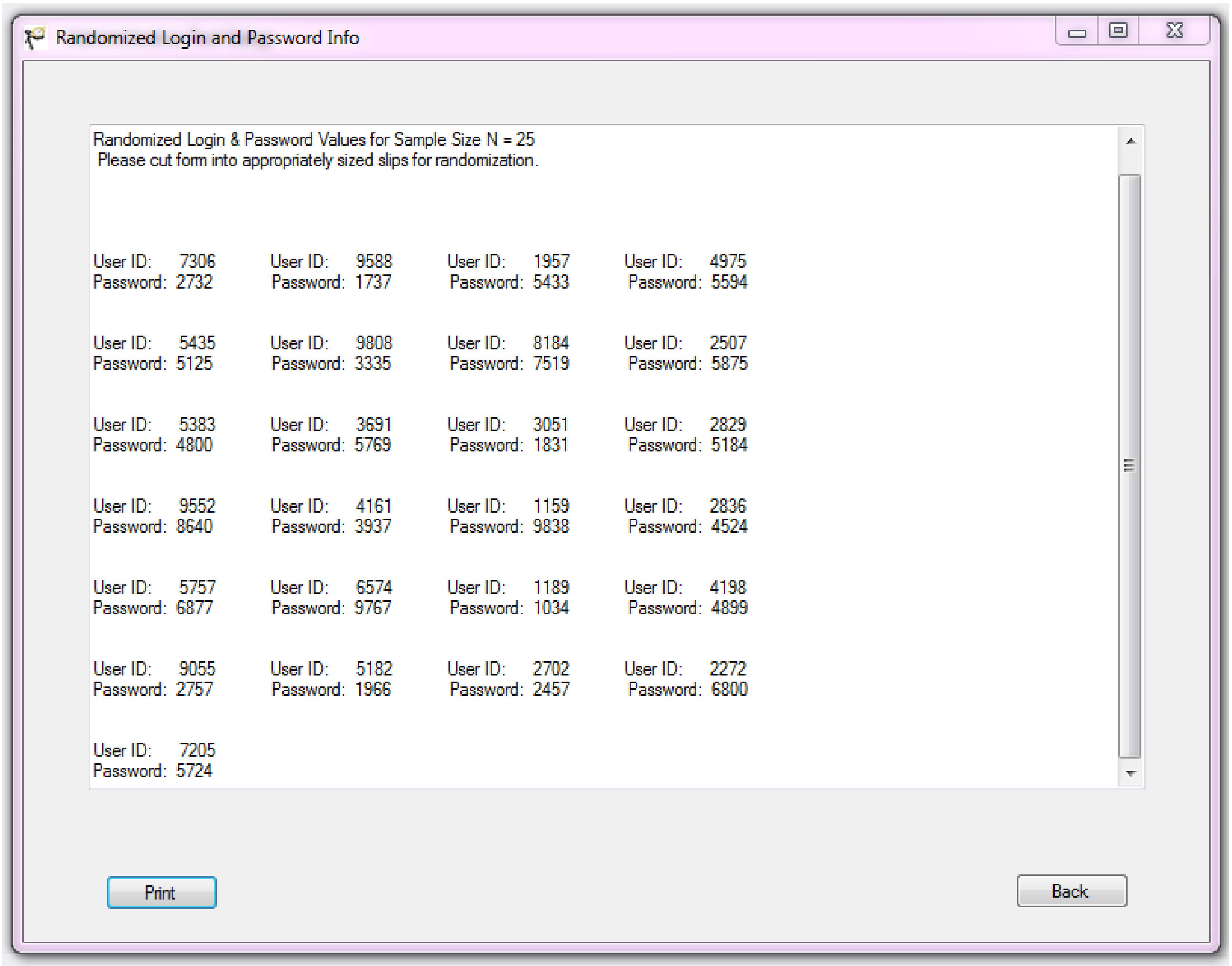


Figure 8 – Generation of user ID and password combinations.

The test proctor prints a hard copy of this screen, cuts the paper into appropriate sized slips and then requests the test subject to pick his/her login information at random. Once the test subject logins to the application, he/she will no longer be identifiable to the study investigators. Although this method is a very low-tech solution, this functionality is important because anonymization of research subjects can be a requirement of institutional review boards (IRBs).

The test proctor will then setup the individual test station for the test subject by entering the user ID and password chosen (Figure 9).

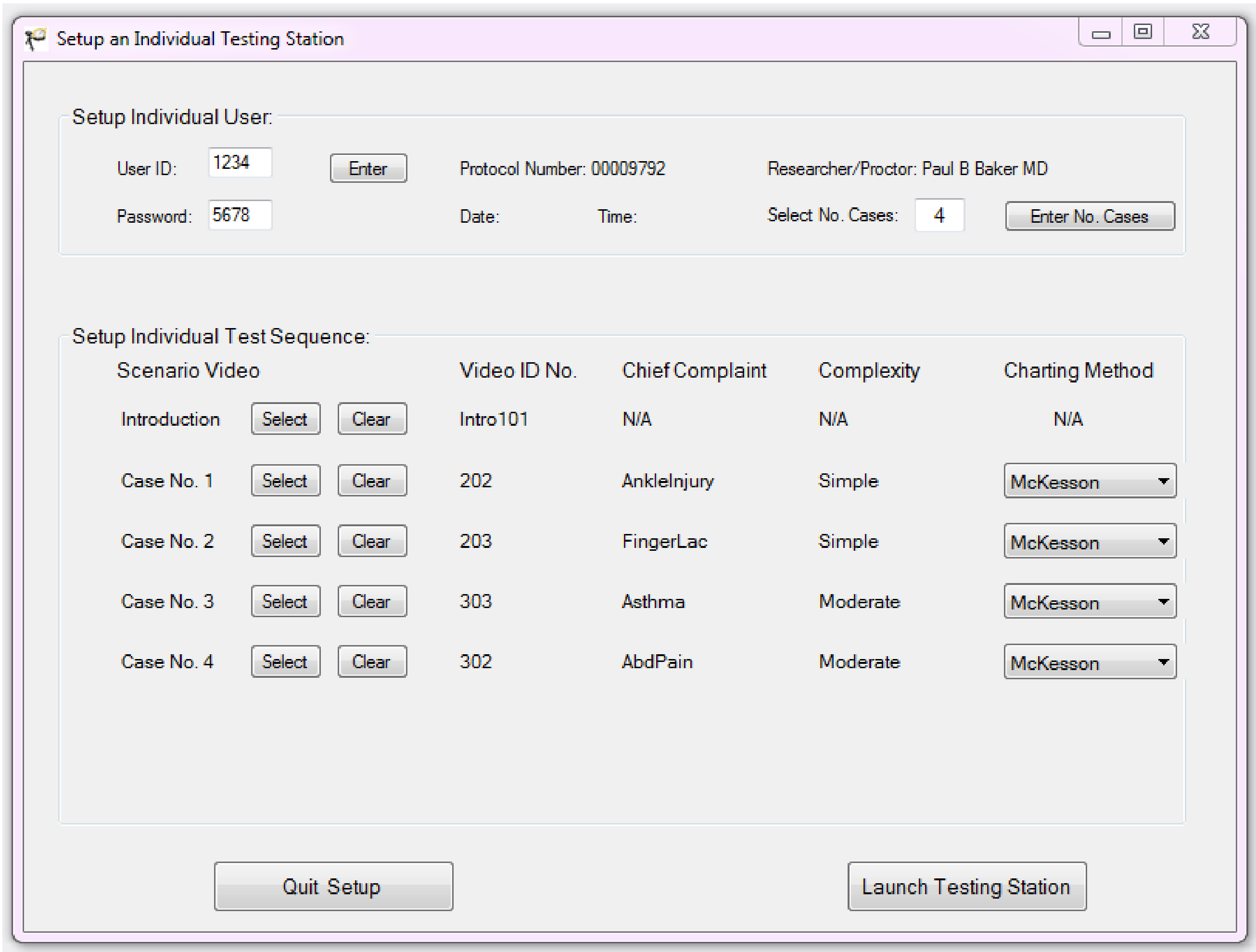


Figure 9 – Setup of an individual testing station.

The test proctor will enter the selected user ID and password from the paper form chosen by the subject. In addition, the proctor will enter the number of cases that will be viewed by the subject during this session. Once entered, the proctor can select the individual test sequence to be viewed by the test subject. The testing palette will show the ID number of the video, the chief complaint and the level of medical complexity of the case scenario video selected. In addition, the software can display which charting method should be used to document the particular case in the sequence. Once the researcher is finished with case selection, he/she can launch the testing station by clicking the appropriate button on the lower right corner of the screen.

The *EHR Test Bench* application is now turned over to the test subject to begin the benchmarking session. The research subject will login into whatever EHR system that is being evaluated. The EHR charting software will be simultaneously running on a desktop computer alongside the smaller computer that is running the benchmarking application. The test user will login to the application by entering his/her user ID and password (Figure 10).

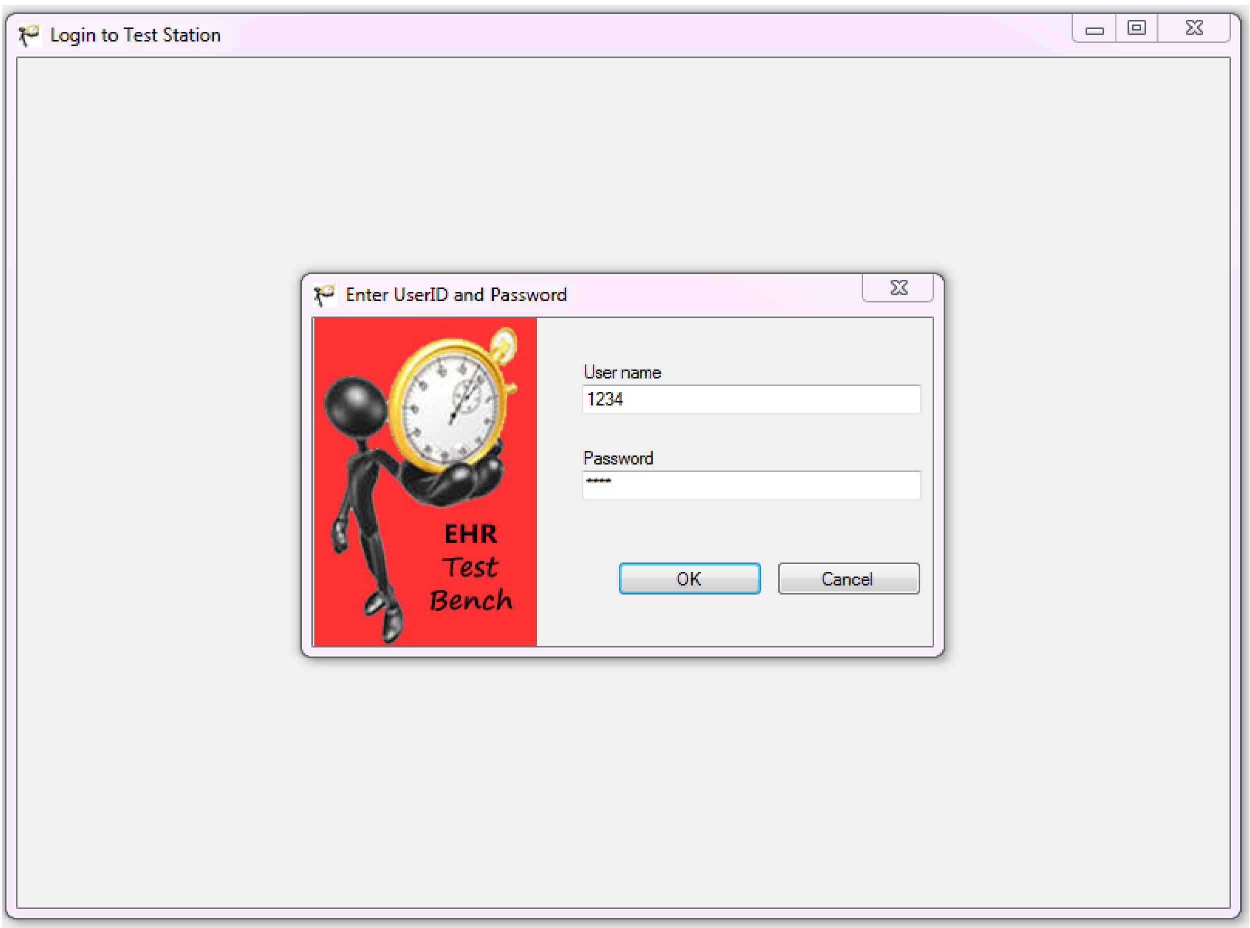


Figure 10 – Login screen for the research subject.

Once the user successfully logins, he/she can adjust the volume and size of the video viewer. The user will then watch a brief introductory video regarding the testing protocol (Figure 11).

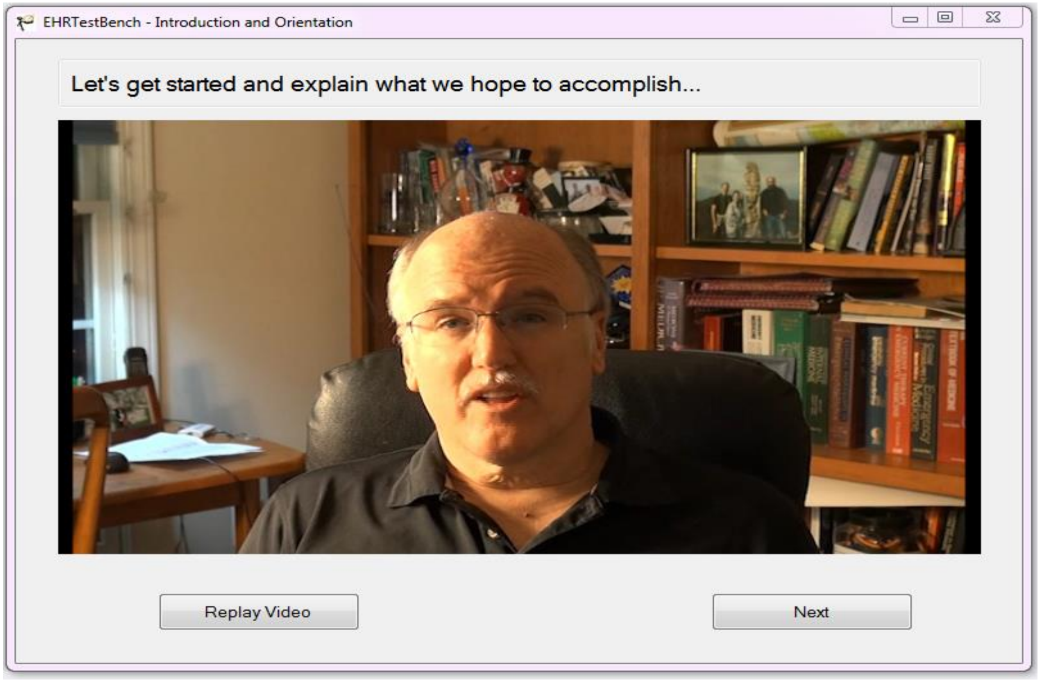


Figure 11 – Introduction and orientation video.

The user will then be prompted to complete a brief questionnaire in Part I (Figure 12).

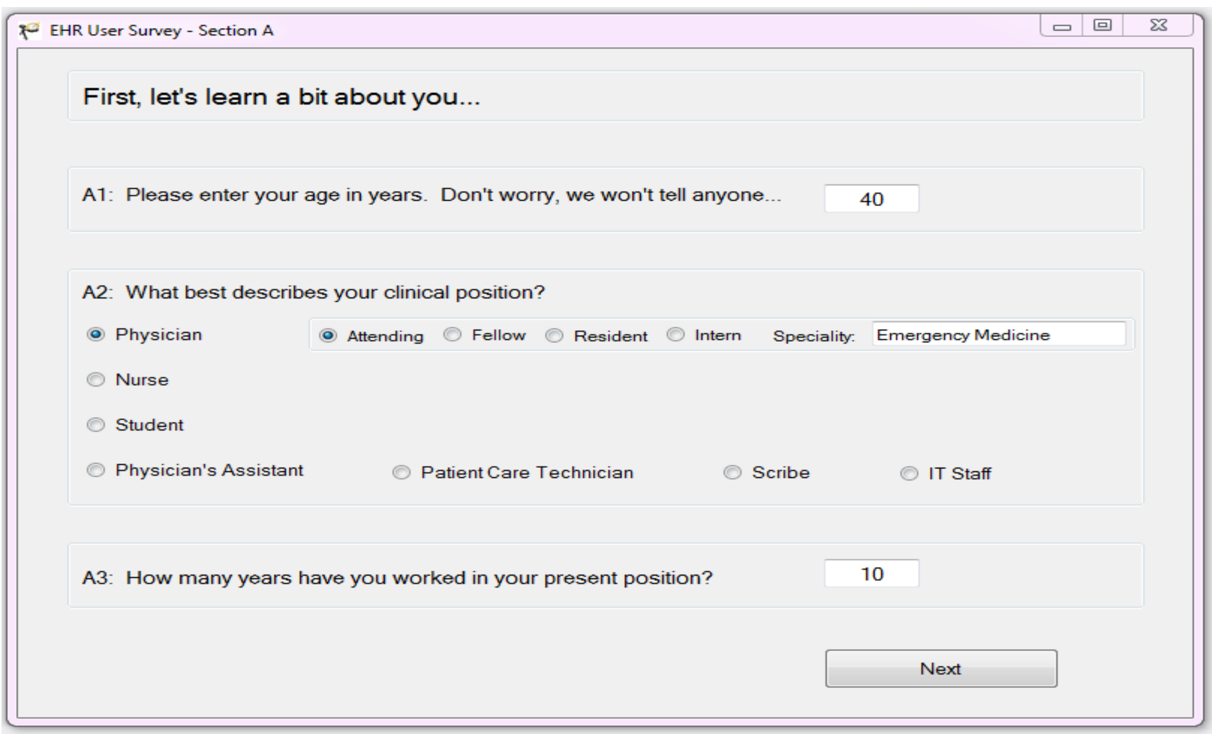


Figure 12 –Part I EHR user survey.

In the next section, Part II, the user will be further oriented to the study and the user interface of the benchmarking application (Figure 13).

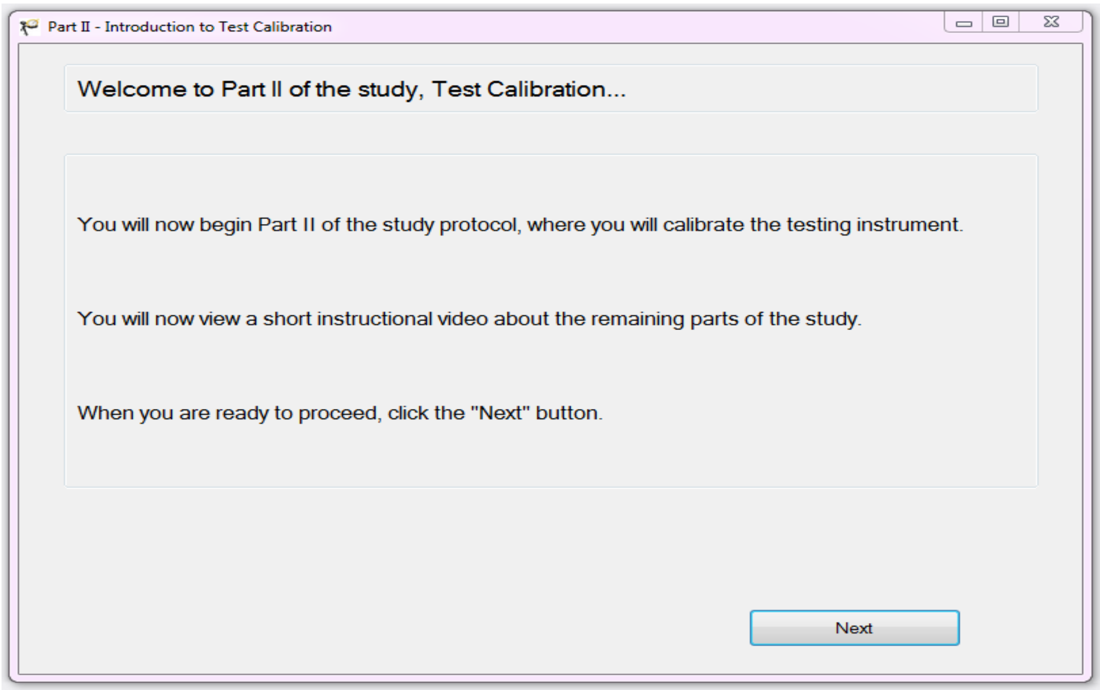


Figure 13 –Part II introduction and test calibration.

The user will watch another video that asks the subject to press various buttons on the screen in response to verbal instructions from the narrator of the video. Although these steps are requested as way to "calibrate" the timing mechanism of the software, actual calibration of the software is not needed. The true purpose of this sequence is to determine the variation among users when asked to repetitively perform sequential screen clicks by recording the time intervals, and thereby collect data regarding the reliability and reproducibility of our testing instrument.

After completing this section of the software, the user should now be familiar with the screen controls and the techniques for starting the video viewer and time stamping when they finish the charting process. The test subject will begin the actual benchmark testing during Part III of the program. It is hoped that test subjects should be able to complete the testing protocol with minimal intervention on the part of the test proctor. The test subject will be presented with the opportunity to begin the first case of the test sequence by pressing a start button on the screen (Figure 14).

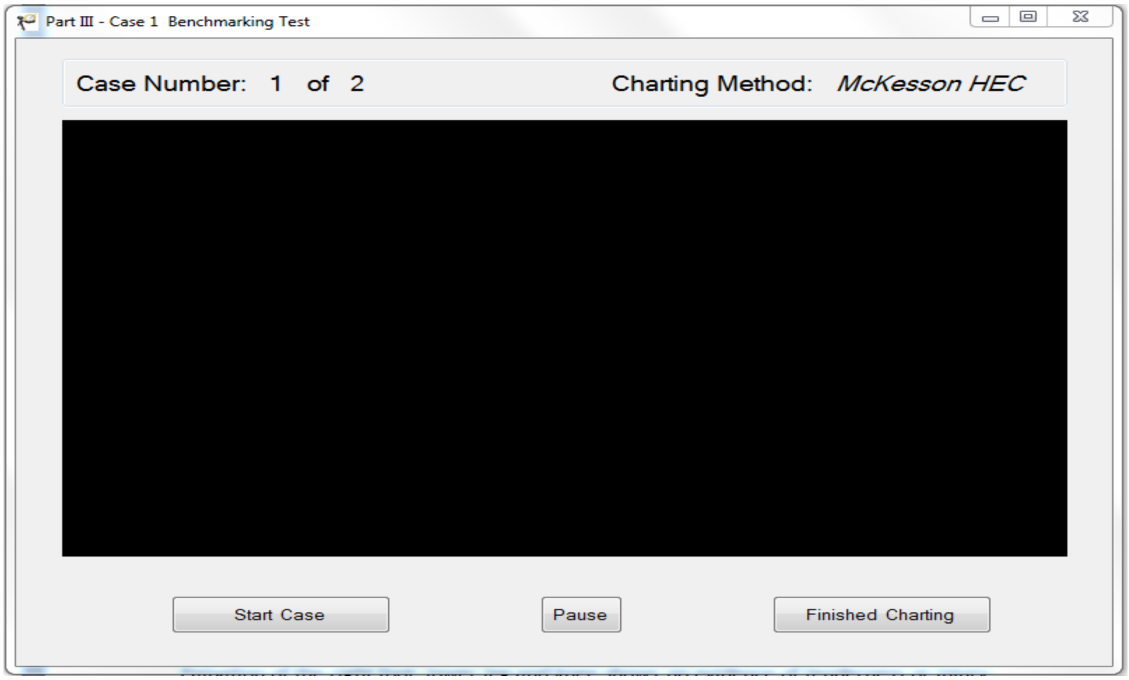


Figure 14 –Part III Benchmarking test prior to starting case.

Pressing the start button timestamps the start of the case and begins the video (Figure 15).

Figure 15 –Part III Benchmarking after clicking start button.

Although users are strongly discouraged from using the pause button, a test subject can temporarily stop the case scenario video if needed in an emergency and the timers will be reset. The test proctor can restart the case video from the beginning by entering his/her administrative password (Figure 16).

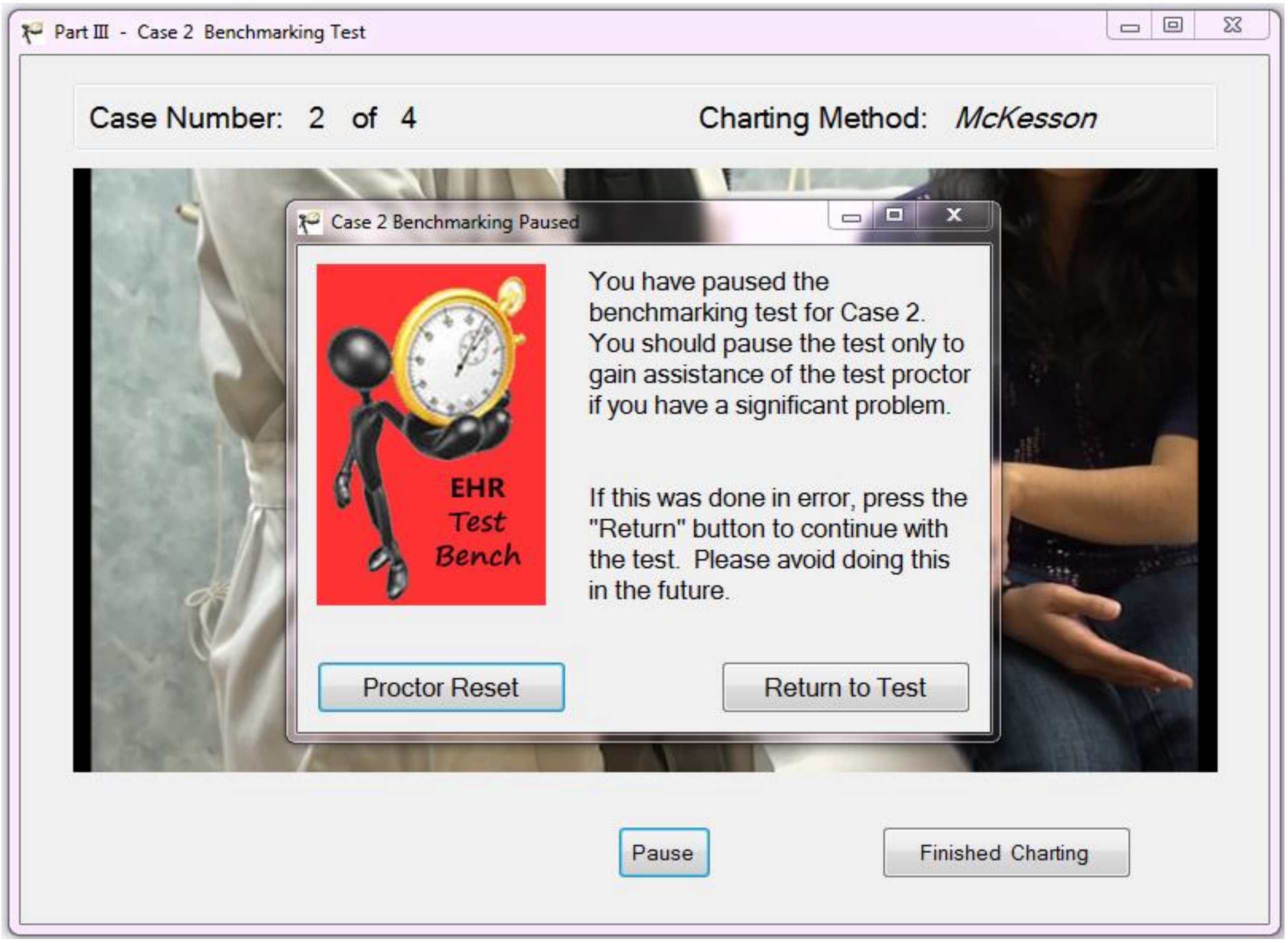


Figure 16 –Pause button can stop video and suspend the timer.

Once the test subject finishes watching the case video, he/she is reminded by a written message on the viewer screen to chart the clinical information in the scenario in the EHR system being benchmarked. When finished charting the case on the desktop EHR, the user will push the finished charting button on *EHR Test Bench* application to time stamp the end of the documentation process. The total documentation time (**DT_{TOT}**) for each case will be written to the data file along with the filename of the clinical scenario.

Before proceeding to the next case, the user will be given a brief (3 to 5 minute) rest period, during which a video clips of medical topics from popular movies will be shown on the viewer (Figure 17).

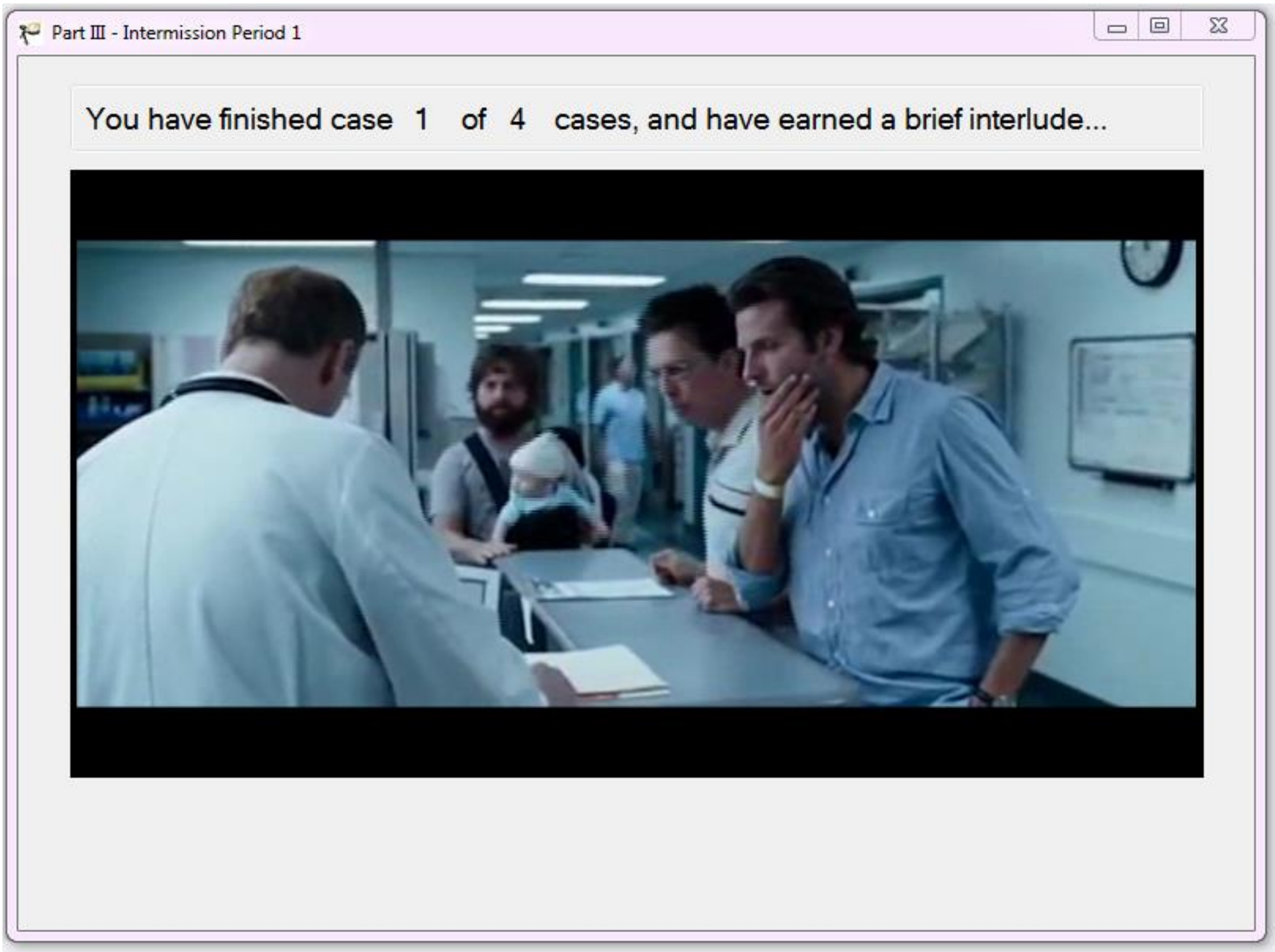


Figure 17 –Intermission period between cases with movie clip.

The next case scenario will follow the intermission in the sequence programmed by the test proctor, up to a maximum of six videos. After completing the final case video, test subjects will watch a short video closing statement thanking them for their participation and asking for feedback. At this point, the involvement of the test subject is finished. The results of the user survey in Part I and the timing measurements from Part II and III of the testing protocol are stored in a data file in simple text format for later analysis. The text file can be directly viewed from the application UI by the test proctor by entering an administrative password (Figure 18).

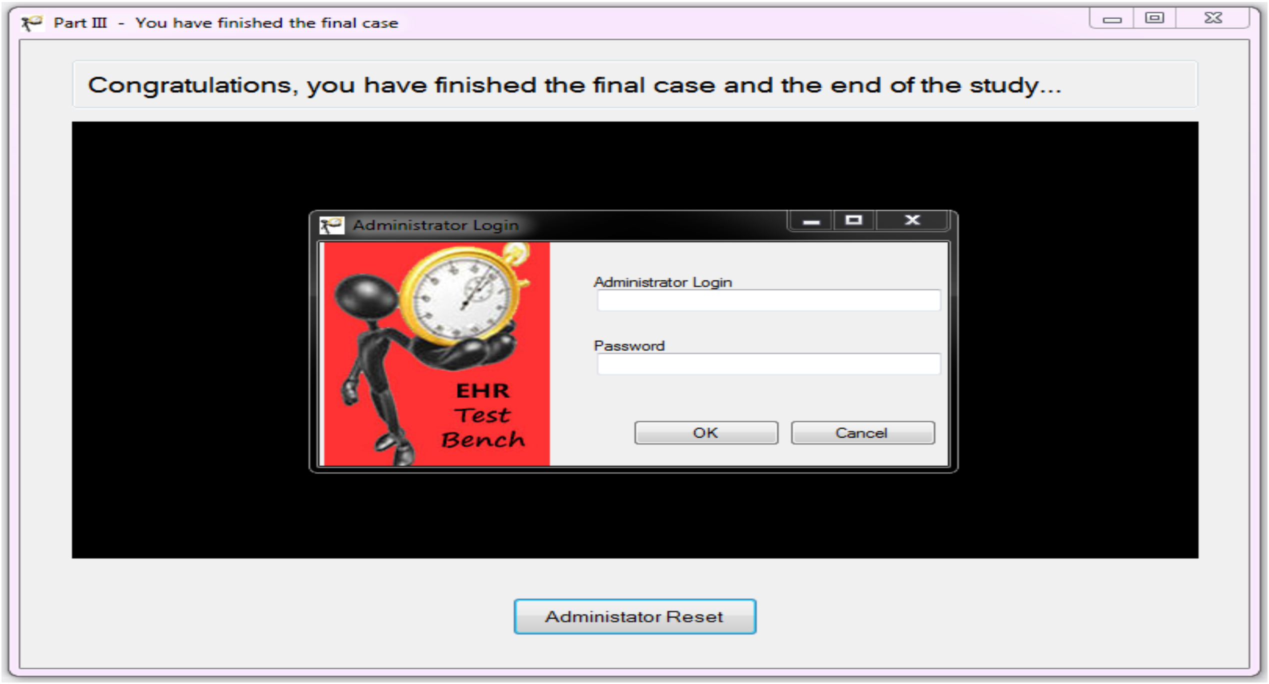


Figure 18 –Administrative login to access study results.

Actual benchmarking results can be viewed onscreen or printed (Figure 19).

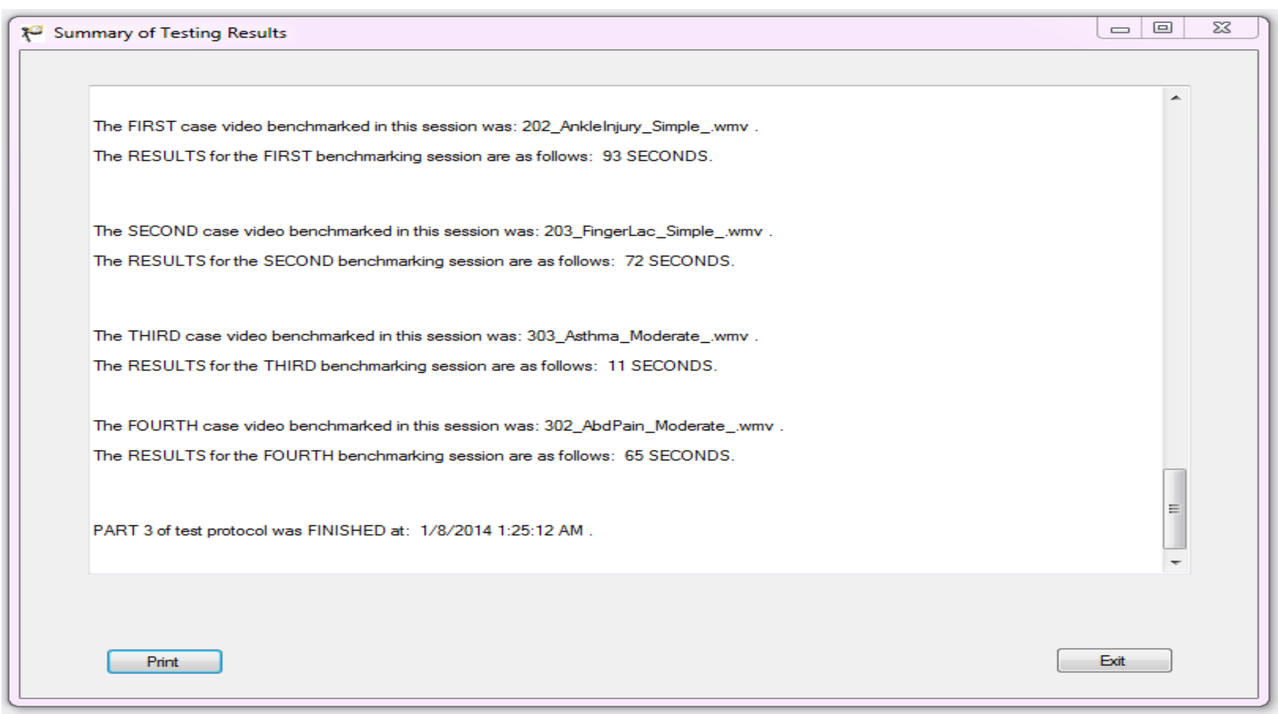


Figure 19 –Summary of testing results.

The programming behind the *EHR Test Bench* application is straightforward. The function of the software is essentially to combine a video viewer with a stopwatch. The main component of the user interface is customized version of the Windows Media Player, which is a standard Active X control object embedded into the application. The standard media player controls (start, stop, pause, etc.) are disabled and replaced with customized buttons that allow the test subject to start the video, view the video once, and then mark the end of the charting process. The goal was to create a simple, easy-to-use interface that allows the user to focus on the main task of watching the scenario and creating a chart of the encounter. When the application is run on the Windows 8.1 operating system, the user can use the touchscreen to enter data and minimize use of the keyboard and mouse. This is an important advantage because the second computer, running the EHR program being studied, will also require use of an additional mouse and keyboard which could be confusing to test subjects.

The algorithm to calculate and record the timing parameters is equally simple. Pressing the “Start Case” button will set one date-time variable equal to the current time, while clicking on “Finished Charting” button will set another date-time variable equal to a slightly later “current time”. The total documentation time ($\mathbf{DT_{TOT}}$) is then calculated by subtracting the two variables, adjusting for any time the case video was paused. Relevant timing data is then written to a text file for later analysis.

Creation of the videos for the case scenarios was less time consuming than expected. Scripts for the patient-provider visits were written to illustrate typical clinical complaints seen in the emergency department. The roles of patients were played by volunteers from the scribe staff working at the hospital. The script for each scenario was only intended as a rough guide, and actors were allowed to improvise during filming as long as the data was clinically consistent with the case and clearly spoken. By not requiring the actors to memorize lines, production time was considerably reduced. Because physical findings such as auscultation of the lungs cannot be conveyed in video format, the clinician needed to “speak aloud” his/her findings to the audience. This convention is already used by ED physicians when using medical scribes to enter clinical data into an EHR and was readily accepted by our test subjects.

The case scenarios were filmed using relatively inexpensive, consumer-quality digital video equipment (Sony Handycam HDR PJ790) fitted on a video tripod. Additional illumination and sound recording equipment was not required, as the lighting levels present in the ED was adequate for filming. The total time to script, rehearse and film the initial six videos for the project was less than eight hours. After filming the case scenarios, the raw video footage was edited using video editing software (Adobe Premiere Pro CC) and converted to the Windows Media Video (.wmv) format, a proprietary video compression codec developed by Microsoft. The videos of the scenarios were edited so that the running time of an individual case averaged between three and five minutes. Further study will be needed to determine the optimal running time of clinical case videos. Scenarios were classified according to complexity, from simple (e.g. ankle sprain) to moderate (e.g. asthma exacerbation) to complex (e.g. altered mental status from a drug overdose). Filename of the video files followed the following naming convention: *ID Number_ChiefComplaint_ComplexityLevel_.wmv*. For example, a scenario involving an ankle sprain would be coded as 202_AnkleInjury_Simple_.wmv. From our experience, it is anticipated that creation of a fairly large library of video case scenarios could be accomplished with motivated volunteers (e.g. medical students) without a significant investment in equipment or resources.

Research Questions

The benchmarking protocol developed for this research study is a preliminary attempt to answer a number of critical questions regarding our study population and its use of electronic health record software. The questions can be divided into four categories:

1.) What are the baseline demographics of the clinicians in our study population? Specifically, we are looking for data regarding age, provider type, specialty, typing skills, and experience with computers in general and with the specific EHR being tested.

2.) How reliable is our testing instrument, the *EHR Test Bench* application, in consistently recording accurate timing data from the benchmarking sessions?

3.) What are the mean performance values for charting efficiency ($\boldsymbol{\eta}_{\mathbf{Chart}}$) and charting accuracy ($\mathbf{A}_{\mathbf{CHT}}$) for the study population? What is the variability for these parameters among the testing group?

4.) How do the clinicians in the study rate the *EHR Test Bench* application and the benchmarking procedures used in the protocol? What are the optimal methods of collecting this type of data from users?

In the limited time available for this capstone study, it is unlikely that a large enough sample of users can be tested for the initial results to achieve statistical significance. However as the benchmarking study continues and a larger sample size is obtained, the researchers will look for any relationships between charting efficiency versus charting accuracy. In addition, results will be analyzed for correlation between EHR performance and variables such as typing skills, provider types and experience levels.

Study Methods

This study is being conducted in the Emergency Department at Holy Cross Hospital in Silver Spring, Maryland, a large community hospital in suburban Washington DC. The emergency department sees a wide variety of general medical, surgical and pediatric complaints, and has a patient volume of approximately 90,000 visits per year. The hospital has used electronic health records in the ED since 1996 and throughout the hospital since 2008.

The pool of test subjects for this protocol consists of employees of Silver Spring Emergency Physicians PC (SSEP), which is contracted to provide emergency medical services at Holy Cross Hospital. The physician group employs approximately 20 adult emergency physicians, 10 pediatric emergency physicians, five nurse practitioners, and five physician assistants, and all are eligible for inclusion in the study. Although residents and medical students rotate through the ED, they are excluded from participation.

The electronic health record study used in this benchmarking study is Horizon Emergency Care (HEC, McKesson Corporation, Version 10.2). The prototype for this EHR was developed and tested in Holy Cross in the late 1990's, eventually licensed by McKesson, and incorporated as their product offering for ED charting. The HEC EHR, in one version or another, was in continuous use in the ED from 1996 until the product was replaced by another EHR system in November 2013. Therefore, all clinicians in the ED are extremely experienced users of the McKesson HEC EHR software.

Benchmarking testing in this protocol is being conducted in the physician offices of SSEP at the hospital, and participation in the study is completely voluntary. The HEC system is setup in the training mode and there is no access to actual patient data or protected health information present in the production mode of the system.

This protocol was reviewed by the Institutional Review Board at the Oregon Health & Science University in January 2014 and met the requirements for exemption and approval in accordance with 45CFR46. 101(b) [2], research involving the use of survey procedures where human subjects cannot be identified.

Testing Protocol

The test setup for the protocol consists of two computers placed on a large table in the physician offices at the ED (Figure 20). The larger desktop computer has the HEC EHR software installed and running in training mode on the hospital network. A smaller computer has the *EHR Test Bench* software installed and placed next to the larger hospital computer. Subjects are provided with pen and paper to take notes, along with headphones to insure privacy. Care must be taken to separate the mice and keyboards of both computers to avoid confusion during the benchmarking session. Users have the option of touchscreen data entry for the *EHR Test Bench* program, which runs on Windows 8.1.

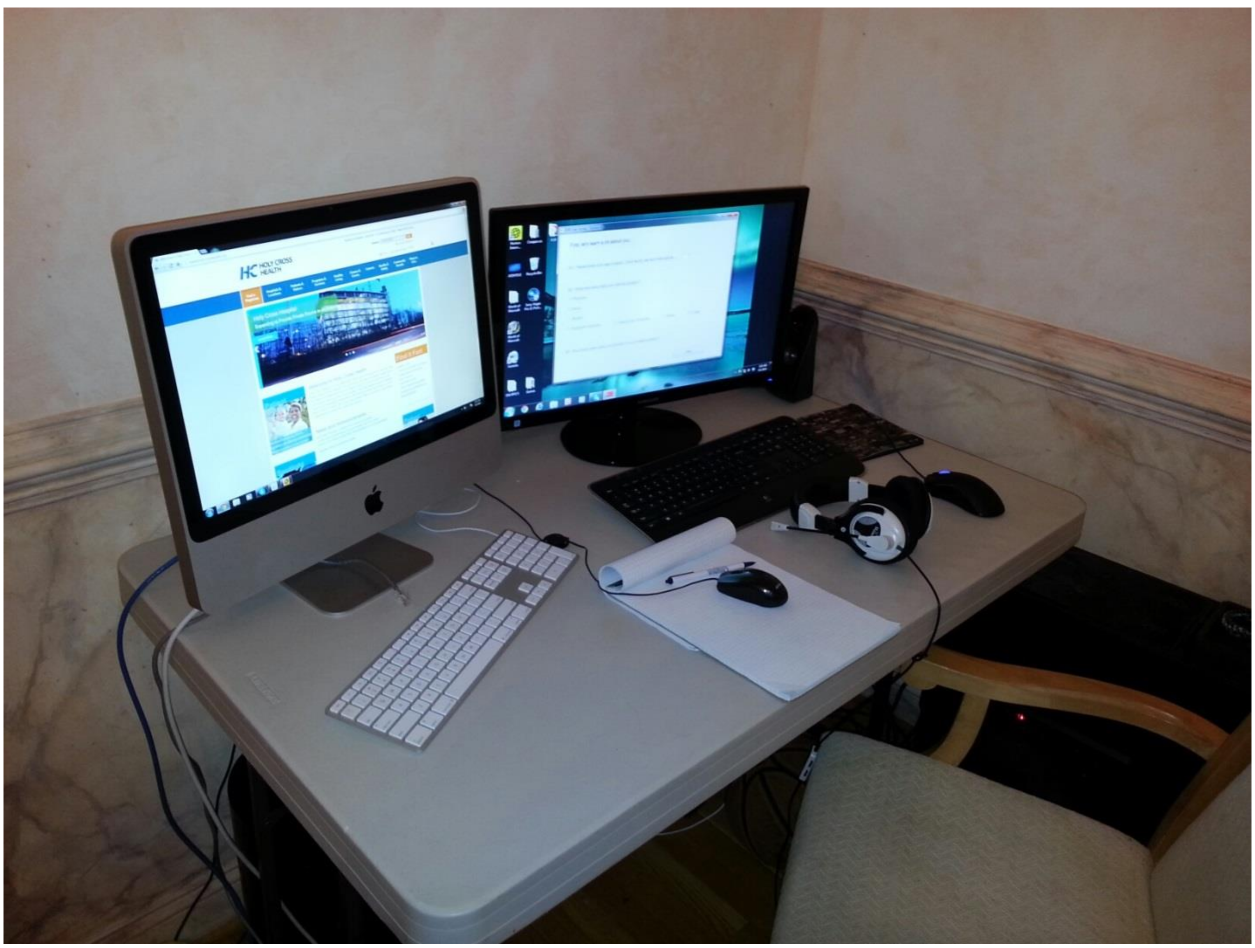


Figure 20 –Photo of test setup using two computers.

The test subject is brought into the room and consented for the study. If the subject agrees to participate, he/she is asked to select a slip of paper containing a user ID and password combination from a large number of slips generated in advance. The user ID is the only identifier used in the protocol and insures anonymization of testing results. Once the subject selects the user ID, the test proctor will create three test patients on the HEC EHR and place these patients in rooms, "ready to be seen." The naming convention for the patients follows the following format: user ID, patient number so that test subject with user ID 2846 would have 2846, First_Patient, 2846, Second_Patient, and 2846, Third_Patient already entered into the EHR and ready to begin charting. The patient ages and chief complaints will also be entered and will match the complaints in the test scenarios.

The test proctor then sets up the individual test station by entering the user ID and password combination and selecting the three case scenario videos that will be shown to the subject. In this early phase study, we have chosen simple and relatively straightforward cases: a teenager with an ankle injury (Case 1), a girl with a finger laceration (Case 2), and a young woman with abdominal pain (Case 3). We have tried to pick clinical scenarios that are within the scope of practice of both adult and pediatric emergency physicians along with physician assistants and nurse practitioners to maximize recruitment from our pool of potential subjects.

After the proctor completes setup of the testing station, he/she answers any questions and checks that both computers are working correctly. The subject is instructed to login to the testing station and begin the protocol. Although the *EHR Test Bench* software is designed to function with minimal intervention, the test proctor remains at a discreet distance but available if assistance is needed.

The application should lead the test subject through the steps of benchmarking protocol, which should last roughly one hour. Part I of the protocol should take approximately ten minutes, during which the user will watch the introductory video and answer several survey questions. Part II will consist of watching an additional orientation and calibration video, which should take an additional ten minutes. Part III consists of watching the three videos and creating three charts using the HEC EHR, a process we estimate will take about thirty to forty minutes.

Once the test subject finishes the protocol, he/she will be asked to complete a brief online survey (*Survey Monkey*) to provide feedback on the *EHR Test Bench* software and our benchmarking procedures. The test proctor will print out hard copies of the three cases and make up a backup copy of the benchmarking data file for later analysis.

The user survey and benchmark timing data from the study will be stored in spreadsheet format and results summarized. We will calculate mean charting time ($\mathbf{DT_{TOT}}$) and charting efficiency (η_{Chart}) along with standard deviation for each of the three cases charted by the group.

Each of the three cases charted during a benchmarking session will be scored using the scoring templates created in advance, and a numerical score for documentation accuracy (A_{CHT}) will be generated for each user. We will then calculate the mean documentation accuracy and standard deviation for each of the three cases charted by the group.

All data will remain anonymized by exclusive use of the user ID as the only identifier and only group results will be reported. All study participants will be furnished, upon request, with a copy of summary results of the testing protocol.

Preliminary Results

Testing began in March 2014 and is continuing at this time. At the time of this preliminary report, a total of ten (10) subjects have completed the protocol, and we hope to enroll an additional ten to fifteen subjects by the end of the study.

After tabulating the data from Part I, the EHR User Survey, we obtained the following summary results:

Question Number	Survey Question	Group Results	Total subjects (N)
A1	**Age in years of Test Subject**	**44.9 (Mean) 44.5 (Median)**	**10**
A2	**Clinical Provider Type**	**Adult EM Attending 7 (70%) Pediatric EM Attending 0 (0%) Physician's Assistant 3 (30%) Nurse Practitioner 0 (0%)**	**10**
A3	**Years in Current Clinical Position**	**15.0 Yrs. (Mean) 16.0 Yrs. (Median)**	**10**
B1	**Self-reported Computer Expertise**	**Novice 4 (40%) Intermediate 4 (40%) Advanced 2 (20%)**	**10**
B2	**Self-reported Typing Expertise**	**Novice 4 (40%) Intermediate 5 (50%) Advanced 1 (10%)**	**10**
C1	**Experience in years with EHR being tested**	**10.7 Yrs. (Mean) 9.5 Yrs. (Median)**	**10**
C2	**Self-reported expertise with EHR being tested**	**Novice 0 (0%) Intermediate 7 (70%) Advanced 3 (30%)**	**10**

Figure 21 –Summary results of EHR User Survey in Part I of protocol.

Part II of the protocol involved collection of three timing measurements for sequences during which the test subject was instructed to press control buttons on user interface screen in response to verbal commands delivered on the video viewer screen. Summary results for the group were as follows:

Sequence Number	Sequence order of button clicks	Mean elapsed time among all subjects μ ± SD (seconds)	Total subjects (N)
1	Start video Finished charting	36.3 ±0.5 seconds	10
2	Start video Finished charting	12.4±0.5 seconds	10
3	Start video Pause/Return to Test Finished charting	41.7±0.9 seconds	10

Figure 22 –Summary results of calibration measurements in Part II of protocol.

Part III of the protocol involved collection of timing and accuracy data for the three cases that were benchmarked. Summary results for timing parameters were as follows:

Case Number	Case Title	Video running time (seconds)	Total documentation time μ ± SD (seconds)	Charting Efficiency μ ± SD (%)	Total subjects (N)
1	Ankle injury	187	546.5±168.9 seconds	37.5%±11.8%	10
2	Finger laceration	192	515.7±120.9 seconds	39.3%±10.0%	10
3	Abdominal pain	262	665.3±203.0 seconds	42.3%±11.9%	10

Figure 23 –Summary results of timing measurements in Part III of protocol.

Summary results for accuracy parameters were as follows:

Case Number	Case Title	Max score obtainable	Mean score for group N ± SD	Mean percent score % ± SD	Total subjects (N)
1	**Ankle injury**	**20**	**16.5±1.8**	**83%±9%**	**10**
2	**Finger laceration**	**22**	**17.1 ±1.9**	**78%±8%**	**10**
3	**Abdominal pain**	**35**	**23.4±6.7**	**67%±19%**	**10**

Figure 24 –Summary results of accuracy measurements in Part III of protocol.

Results from the software evaluation survey for the *EHR Test Bench* benchmarking application are listed in detail in Appendix E. Highlights of the survey include that 70% of study participants felt that this type of benchmarking testing was an accurate method of evaluating EHR software and 50 % would take part in further benchmarking sessions if available over the Internet. Overall, 100% of respondents found the case videos to be “realistic”, and 70% would like to see more complex case scenarios presented in future testing. A majority of users (90%) thought that the ideal number of videos in a session was three cases, while opinions regarding the optimal duration of a testing session were variable, ranging from 15 minutes (30%) to a half hour (40%) to a maximum of 45 minutes (30%). From a software designer’s perspective, it was gratifying to find that 100% of users found the application to be “very user-friendly”. However, 10% of subjects admitted asking the test proctor for assistance to complete the protocol, suggesting that further refinements and product testing will be necessary before *EHR Test Bench* can truly be considered as a standalone application.

Data Analysis

At this point in our study we are limited in the amount of data analysis that can be performed, given the small sample size thus far in the study. However, we can already see some trends emerging from the results for the benchmarks obtained from the three clinical scenarios presented to the test subjects. We look first at the time taken to document each of the three clinical encounters. We have previously defined the charting efficiency as the ratio of the encounter time (running time of the video) over the total documentation time (running time of video plus time to finish charting) expressed as $\boldsymbol{\eta_{Chart} = T_{TOT} / DT_{TOT} \times 100}$. A graph of charting efficiency for each of the cases shows the following distributions:

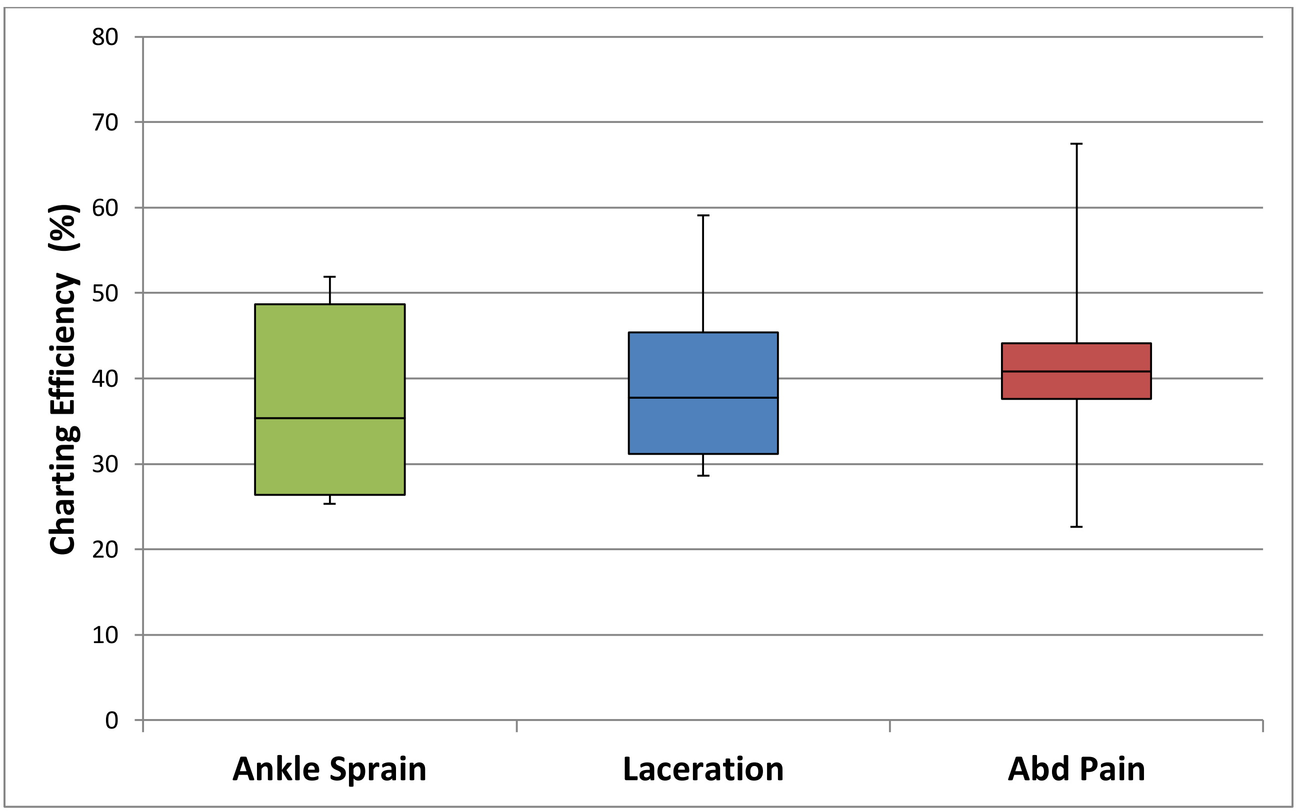


Figure 25 – Box plot of charting efficiency (η_{Chart}) for clinical cases.

It can be readily appreciated from the box plot that there is a fairly broad distribution of charting efficiency among the clinicians when documenting all three clinical scenarios, even though all test subjects were very experienced daily users of this EHR documentation system.

As might be expected, there are outliers on both extremes: highly efficient “charters” who approach 60% efficiency while other, less efficient individuals document in the 20 to 30% range. It is important to note that even for these relatively uncomplicated clinical scenarios, three quarters of the clinicians studied were measured at less than 50% charting efficiency, which suggests that a typical provider using this EHR spends more time in charting activities than in interacting with actual patients.

Turning to charting accuracy, we see an equally wide range of performance among our test subjects. A graph of the documentation accuracy scores ($\mathbf{A}_{CHT}$) for each of the three cases shows the following distributions:

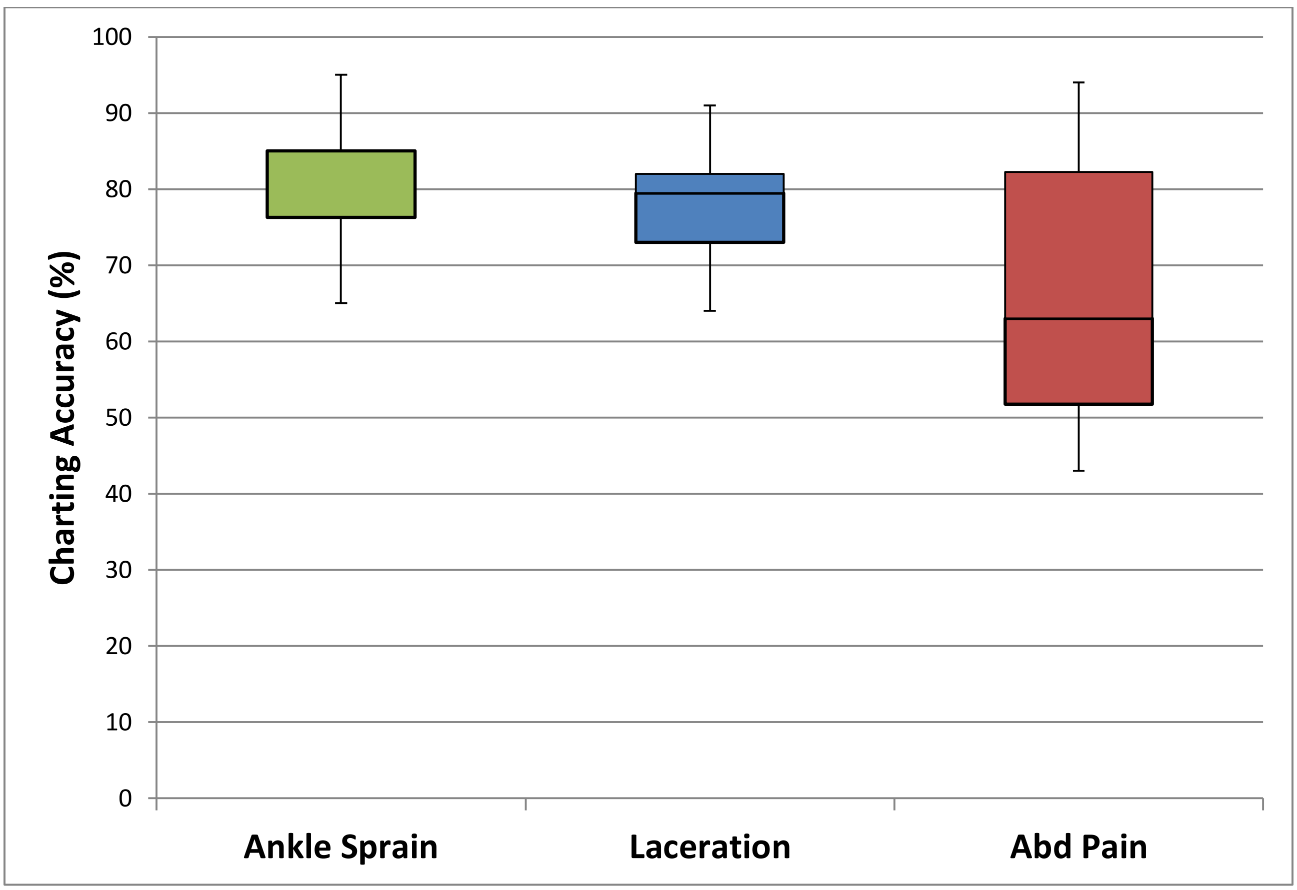


Figure 26 – Box plot of charting accuracy ($\mathbf{A}_{CHT}$) for clinical cases

Likewise, there is a broad range of scores for clinical accuracy for test subjects on all three cases. Again there are outliers in both extremes: highly accurate “charters” who capture in excess of 90% of clinical data points, while others who miss up close to half of those same data points. As one might expect, there was a larger range of accuracy scores for the most complicated clinical case scenario in our study, a patient with abdominal pain. This case video had the longest running time of our series and also had the highest number of clinical data points that could be captured or omitted. A similar larger spread for charting efficiency was noted for this same scenario when compared to the two less complicated cases, ankle sprain and a simple laceration.

Are the fastest, most "efficient" charters the same individuals who are the most accurate? Is documentation accuracy sacrificed for documentation speed? Unfortunately, our small sample size at this point will not allow analysis of the correlation between charting speeds versus charting accuracy. Likewise, we cannot yet the role of typing speed, provider type or age, and computer expertise in benchmarking performance. We hope to conduct these further analyses once we have enrolled more test subjects.

We have attempted to measure the reliability of our testing paradigm that uses a simple software application and user interface to mark the beginning and end of the charting process to calculate documentation times. We analyzed the timing data (a series of sequential button clicks in response to verbal commands, figure 22) in order to quantitate the internal consistency reliability of our testing instrument. Using our sample size (N=10) we calculated a Cronbach's alpha (α) of 0.73 for our three timing measurements. Most authorities believe that a Cronbach's $\alpha > 0.70$ is an adequate level of reliability for comparison studies.[42] As our sample size increases, we expect this value to rise according to the Spearman-Brown prophecy.

Discussion

It is difficult to draw definitive conclusions from this preliminary study with limited scope and small sample size, especially in a protocol whose stated goal is to obtain rigorous quantitative benchmarking data about EHR software systems. Nevertheless, we can make some noteworthy observations and identify interesting trends in the results obtained thus far.

First, and perhaps most importantly, this protocol demonstrates that it is feasible to collect comparative data on EHR performance and efficiency in a systematic, standardized, and reproducible fashion. Furthermore, the process of benchmarking an EHR system can be performed relatively quickly and inexpensively---elaborate usability laboratories and costly time and motion studies are not needed to obtain useful and clinically relevant results.

Second, much of the data collection and analysis required for benchmarking can be automated using simple software solutions. Although rudimentary, the *EHR Test Bench* software was successfully used to deliver both clinical content and collect timing data during the study. The application was readily accepted by our test subjects, and serves as a proof of concept and a template for further development of more elaborate and fully featured benchmarking software.

Finally, the circumstances of our study setting and population are unique. The clinicians in our test population were from a single specialty (emergency medicine) and all were skilled system users who, on average, have been using the same EHR at the same hospital for over a decade. Therefore, we were able to study “highly experienced” users working with a “fully mature” EHR ---in many ways, an ideal environment in which to conduct our preliminary benchmarking measurements.

Despite this homogeneity in our test population, we still observed considerable variability in the charting speed or efficiency of clinicians when documenting all three case scenarios. The “fastest” clinicians were able to complete their charting efforts almost twice as quickly as the “slowest” clinicians on all three cases, even though all users had to complete the same charting templates in the EHR software, using the same point-and-click and keyboard entry devices.

In some respects, it is not unexpected to see such variation. There is considerable variation in clinician efficiency in performance of most clinical tasks: parameters such as patients seen per hour, lab studies ordered, and time spent interacting with patients vary greatly among providers. Therefore, we would anticipate that documentation speed would likewise show a range of performance scores derived from any cohort of clinicians. It remains to be determined if the variation in charting efficiency follows a normal distribution, which would aid in statistical analyses of test results.

One goal of quantitative benchmarking would be to report, hopefully with some degree of statistical confidence, the mean charting efficiency scores of competing EHR systems in documenting a history and physical exam from a standardized suite of clinical case scenarios. Having numerical scores validated and published by an independent and unbiased testing organization would allow clinicians and health care organizations make better informed purchasing decisions and demand meaningful improvements in existing system performance from EHR vendors. Although the overall usability and utility of a large software application cannot be reduced to numerical scores, the current lack of these metrics is a major shortcoming in both the marketplace and in informatics research.

Obviously, charting speed is only part of the overall EHR performance equation. No matter how quickly clinical information can be documented, entering inaccurate, incomplete or incomprehensible data into the medical record is always a risk with electronic charting systems. Therefore benchmarking data on charting efficiency must always be paired with and interpreted in the context of charting accuracy.

Our test subjects showed similar variability in scores for charting accuracy for each of the three case scenarios. For the two uncomplicated cases, an ankle sprain and a finger laceration, mean charting accuracy ($\mathbf{A}_{\mathbf{CHT}}$) was in the 80% range, while mean accuracy for a moderately complex case of acute appendicitis dropped to below 70%. Once again, there was a twofold difference in the charting accuracy between the most and least “accurate” clinicians on the abdominal pain scenario. Are the less efficient charters slower because they are taking more time to create a more accurate chart? At this point in our investigation, we do not know. We look forward to entering more subjects in our study, in order to search for just this type of correlation between charting speed and accuracy.

The charting accuracy score ($\mathbf{A}_{\mathbf{CHT}}$) in our protocol must be interpreted with caution and with an awareness of its limitations. By its nature, creating a scoring “template” is a somewhat arbitrary and simplistic measure of charting accuracy. What one clinician believes is essential information may be considered superfluous by another. In addition, scoring of the completed charts was performed by a single individual, the author, while using a panel of judges would have been more rigorous methodology.

Even more importantly, our scoring algorithm made no attempt to adjust for or penalize the clinician for recording clinical data in the chart that was not explicitly present or mentioned in the clinical encounter. Documentation “filler” or the inclusion of superfluous boilerplate text is a very significant issue in the charts authored by clinicians using template driven EHR systems. In an attempt to quickly create a legally defensible chart and maximize reimbursement for care, providers are generating bloated documentation that may bear little resemblance to what actually transpired during the clinical encounter. Apart from the dubious validity of this charting approach, these template driven charts are extremely difficult to read and impart little useful information to other medical providers.

The first priority of the medical record is to summarize clinical data so that it is readily available to other providers who will continue care of the patient. In this respect, quantity does not necessarily translate to quality, and having a completely accurate chart does not guarantee an easily readable or clinical useful one. A challenge in the future will be to refine our scoring algorithm for charting accuracy to better measure charting quality---taking into account less easily quantifiable parameters such as readability and brevity of the clinical document.

We readily acknowledge the limitations of the current study. First, we are constrained by a small sample size at this point and are underpowered to make statistically significant and clinically relevant associations from our data. We plan on enrolling, at a minimum, another ten clinicians over the next month which should increase the sample size to greater than twenty subjects. Although our data collection time has been reduced by using the *EHR Test Bench* application, it still takes at least 90 minutes to consent and complete the test protocol for each test subject along with an additional hour to score the charting documents. In addition, emergency department providers work rotating shifts, and scheduling testing sessions can be challenging. With a single investigator, the data collection process has been far more labor intensive than anticipated.

Another valid criticism of the study is that the testing paradigm of having a clinician watch a video of a simulated medical case and record the details is not representative of what that same provider would do if he or she was taking a history and physical from an actual patient. In reality, we are testing the physician as a scribe rather than as a physician. However, one could argue that with the advent of electronic health records, clinicians have assumed increasing responsibilities of the scribe---spending much of the workday in front of a keyboard and computer. In the words of a colleague, the physician has become a low-level data entry clerk. If this is the functionality we are actually measuring, such data is still a valuable metric. Furthermore, we are studying the performance of the combination of physician and computer, the underlying premise of Friedman's theorem, as opposed to the physician alone. Taken in this context, we believe the focus on the physician as scribe is a valid construct.

A final shortcoming of our experimental design is that we have chosen to benchmark a relatively older and not particularly highly rated EHR system, Horizon Emergency Care from McKesson. The system was chosen for two reasons. First, the system was readily available to the investigator and could be easily accessed in training or non-production mode. Second and more importantly, the clinicians at the study hospital had many years of experience with the EHR and we felt this familiarity would make for a more homogenous study population. While the emergency department at the study hospital has switched to a newer EHR (Cerner First Net), the physicians had less than six months experience with the replacement system and were still learning the application at the time this study began.

In the future, it is hoped that EHR software evaluations such as described in this paper will be expanded in both scope and sophistication. It is clear that a much greater commitment of time and resources will needed if large scale benchmarking studies involving multiple EHR systems and practice settings are going to be successful. Ideally, an academic medical institution or residency program with access to clinical expertise, software development teams, and volunteer actors could readily sponsor these investigations. Alternatively, either non-profit (e.g. CCHIT) or for-profit entities (e.g. Gartner, KLAS) could conduct similar studies. A "consumer guide" for the electronic health record marketplace is sorely needed.

One early and promising approach would be to shift the *EHR Test Bench* software from a standalone desktop program to a web-based application. Placing the case scenario videos and timing functionality on a web server would allow for benchmarking on a far larger scale with hundreds of test subjects and multiple EHR systems. Although many software vendors disapprove or even prohibit such "unauthorized" testing of their products, there would be little to prevent enterprising end-users from remotely accessing their EHR system from home and conducting a benchmarking session over the internet. The potential of such clandestine testing is intriguing.

Benchmark testing does not have to be limited to a clinician taking a history and physical in simple or uncomplicated cases. Testing could be expanded to more complex scenarios (e.g. an acute stroke patient receiving thrombolytics) or vague clinical presentations (e.g. the weak and dizzy patient). Other types of time-consuming clinical data entry tasks, such as order entry or medication reconciliation, could likewise be studied in a systematic, quantitative manner. Most importantly, using benchmark testing in a prospective fashion could aid system developers in designing and optimizing EHR software before these systems reach the marketplace.

In conclusion, the author hopes that benchmark testing will establish new standards of excellence for EHR software in the clinical workplace. By careful and disciplined study, we can ensure that Friedman's Theorem is unequivocally true. Patients and clinicians deserve no less.

Appendix A

Summary Description of Clinical Scenarios

Case 1: A 16 year old male presents one hour after injuring his right ankle playing basketball. The patient reports pain in the ankle and difficulty walking. The patient is in good health with the exception of mild asthma. He has had a previous left ankle fracture that required surgery. The patient is a high school student.

His physical exam shows moderate tenderness and swelling of the right lateral malleolus. There is no tenderness on examination of the right medial malleolus and the Achilles tendon is intact. Palpation of the right foot, lower leg and knee shows no evidence of tenderness or injury. Distal pulses of the foot are presents and neurovascular status is intact.

Radiographic studies of the right ankle show no fracture and the patient is diagnosed with a right ankle sprain. Treatment plan includes a stirrup splint (Air Cast), crutches and limited weight bearing for one week, and follow-up in one week.

Case 2: A 22 year old female presents one hour after cutting her left index finger with a knife while trying to cut a bagel. The patient is healthy except for a history of hypertension. Her last tetanus shot was the year previous.

Her physical exam shows her to be in no distress. There is a single linear 1.5 cm laceration on the dorsum of the proximal phalanx of her left second digit. There is full range of motion including flexion and extension against resistance and tendon function appears to be intact. Distal sensation and capillary refill are normal and neurovascular status is intact.

The digit is anesthetized using a digital block with 1% lidocaine and the wound is explored. There is no evidence of a tendon laceration and the wound is closed with two interrupted 5-0 nylon sutures. Treatment plan includes splinting the finger for the next three days with suture removal in ten days.

Case 3: A 26 year old female presents with a two day history of abdominal pain that has worsened over the past twelve hours. The pain began in the epigastric region and has migrated to the right lower quadrant. Patient is anorectic and nauseated but is not vomiting. There is no diarrhea and her last bowel movement was yesterday. Her last menstrual period was one week prior. She denies she is pregnant and uses birth control pills for contraception.

Her past medical history is significant for migraines and peptic ulcer disease. Her past surgical history includes a cholecystectomy 6 month prior to this admission and a prior caesarian section. She denies tobacco or drug use and has not consumed alcohol since college.

Her physical exam shows her to be in moderate distress. HEENT shows no evidence of jaundice but dry mucous membranes suggesting dehydration. Pulmonary and cardiac exams are normal. Abdominal exam shows hypoactive bowel sounds and marked tenderness in the right lower quadrant. There is guarding and rebound tenderness present in the right lower quadrant.

The patient is given intravenous hydromorphone (Dilaudid) with improvement in her pain level. A CT scan of the abdomen is performed and shows an inflamed appendix consistent with acute appendicitis. The patient is admitted to the hospital under the care of the surgeon on call, Dr. Smith. She is advised that she will be taken to the OR tonight for an appendectomy.

Appendix B

Sample Clinical Scenario Script

Video: Ankle Injury

Filename: 201_AnkleInjury_Simple_.wmv

Doc: What's brings you to the ER today?

Patient: I injured my right ankle playing basketball. I went up for a jump shot and came down on it wrong.

Doc: When did it happen?

Patient: Just about an hour ago. It didn't hurt at first, but it is really killing me now.

Doc: Can you bear weight on it?

Patient: Yeah, I can walk on it but I'm limping pretty badly.

Doc: Did you get hurt anywhere else?

Patient: No, just the ankle.

Doc: Have you injured this ankle before?

Patient: No, but I did fracture my left ankle when I was ten years old and they had to operate on it and put some pins in. I was skateboarding.

Doc: Guys like you keep us in business.

Patient: Thanks.

Doc: Are you pretty healthy otherwise?

Patient: Just some asthma, mostly in the spring.

Doc: Do you take any medications?

Patient: I use an Albuterol inhaler, but only when I get symptoms.

Doc: Any allergies?

Patient: My mom says I am allergic to penicillin, got a rash when I was a baby.

Doc: What grade are you in?

Patient: I'm going to be a junior this year. I was hoping to make the varsity basketball team this year if this ankle holds up.

Doc: Hopefully this will just be a temporary setback. Let's take a look at your ankle.

Patient: OK.

Doc: You look like you have a lot of swelling on the outside of your right ankle. Do you hurt right here? (Doc pushes on the lateral malleolus).

Patient: Yes.

Doc: You are very tender over the lateral malleolus of the right ankle with a moderate amount of swelling.

Doc: Let's check out the rest of your foot and ankle. Do you have any tenderness here, here and here? No---that great. The good news is that you have no tenderness in the rest of the foot and ankle. There is no tenderness in your knee or in your lower leg.

Doc: I'm going to pull on your ankle a bit and it may hurt. Good, your ankle seems to be stable. Let's check your pulse. It's great. Wiggle your toes. I would say that you are neurovascularly intact.

Doc: Let's take a look at your x-rays. (He looks at x-rays.) Good news, no fracture.

Doc: We are going to put a splint on your ankle, get you some crutches. You need to give this a few days. You should follow up with your team trainer in the next week.

Appendix C

Sample Scoring Sheet

Video Name: Ankle Injury

Filename: 202_AnkleInjury_Simple_.wmv

Date created: 11/16/2013

Total Video Running Time: 3:07 (187 seconds)

User ID: **Reviewer:**

Date/Time Session: **Review Date:**

Case Video Sequence No: 1

Total Data Points Available:

Total Data Points Noted Correctly:

Total Data Points Noted In Error:

Data Points Absent:

User Total Raw Score:

User Percentage Score:

User Total Charting Time:

Data Points Scoring:

Data Point	Present (+1 pt.)	Absent (0 pt.)	Erroneous (-1 pt.)
1. CC: Ankle injury			

Data Points Scoring:

	Data Point	Present (+1 pt.)	Absent (0 pt.)	Erroneous (-1 pt.)
2.	HPI: Location R ankle			
3.	HPI: Playing basketball			
4.	HPI: 1 hour ago			
5.	HPI: Mechanism – twisted			
6.	PMHx: Prior L ankle FX			
7.	PMHx: Asthma			
8.	PE: Location R ankle			
9.	PE: Swelling R lateral malleolus			
10.	PE: Tenderness R lateral malleolus			
11.	PE: Non tender R medial malleolus			
12:	PE: Achilles tendon – intact			

Data Points Scoring:

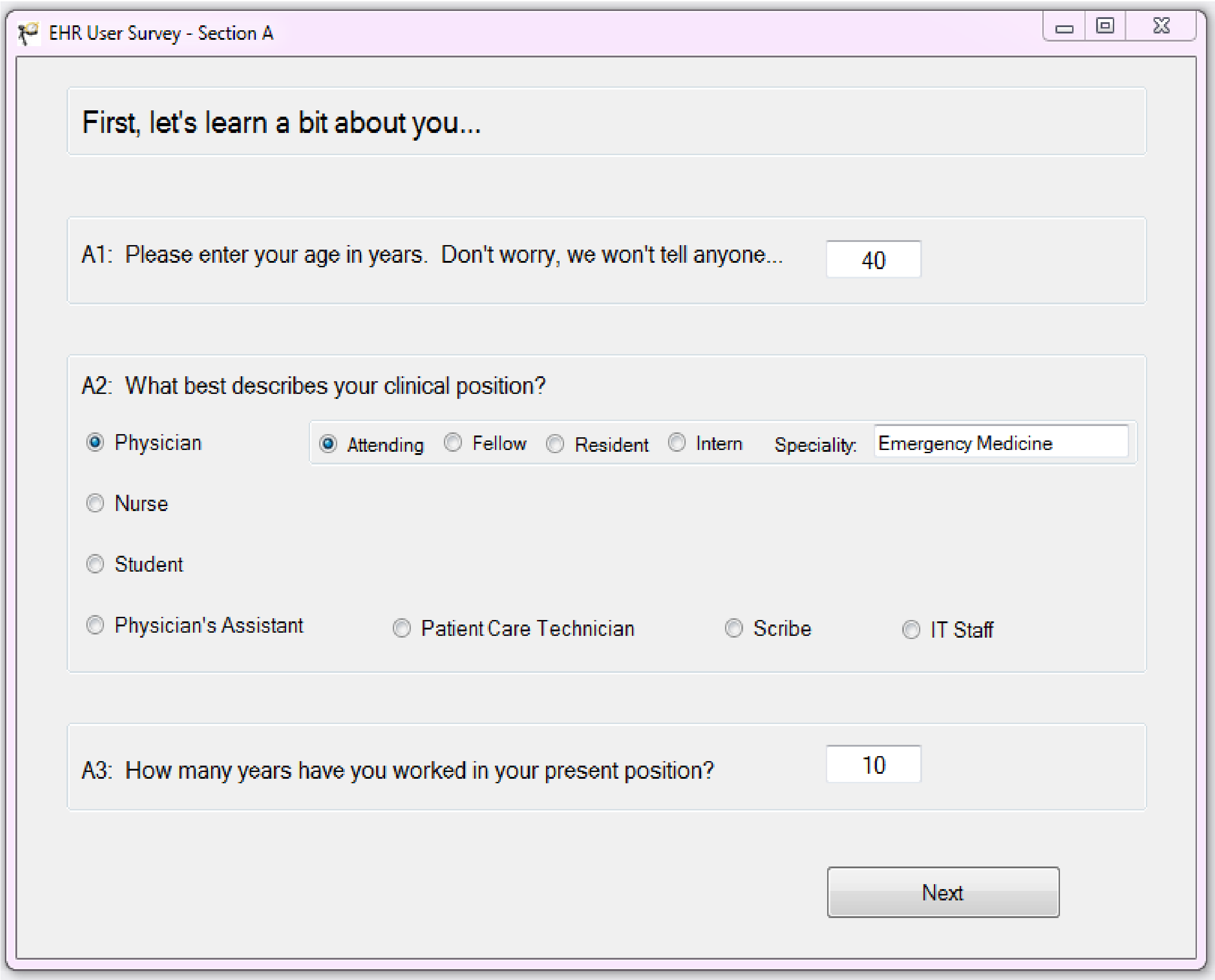
	Data Point	Present (+1 pt.)	Absent (0 pt.)	Erroneous (-1 pt.)
13:	PE: Foot – non-tender			
14:	PE: Knee – non-tender			
15:	PE: Calf - non-tender			
16:	PE: Neurovascular status intact			
17:	Results: X-ray- no FX			
18:	Dx: Ankle sprain			
19:	Treatment: Splint/Air cast			
20:	Treatment: Crutches			

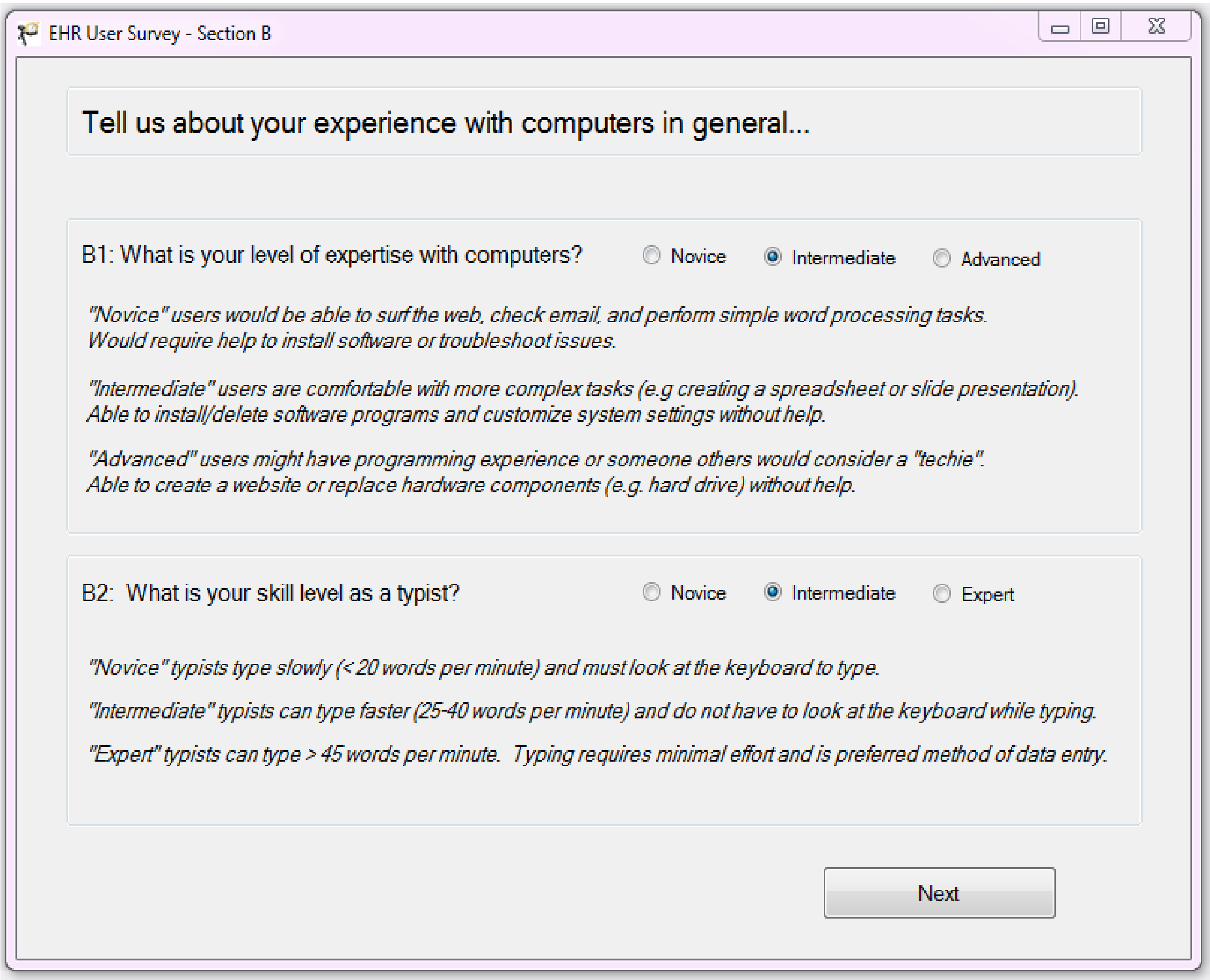
Total Points: ____/_____

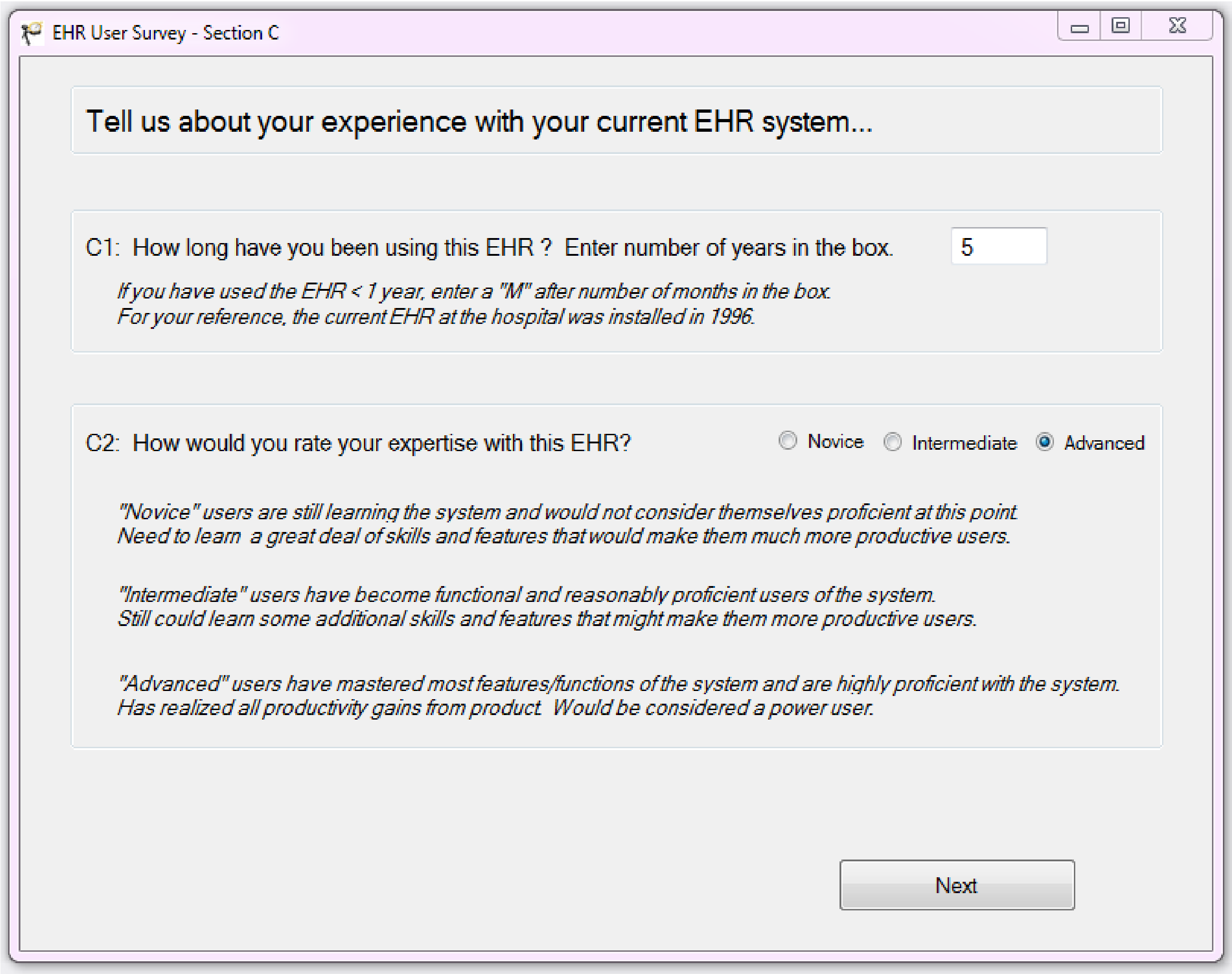
Appendix D

User Survey Questions

From Part I




EHR User Survey - Section B
Tell us about your experience with computers in general...
B1: What is your level of expertise with computers?
Novice
Intermediate
Advanced
"Novice" users would be able to surf the web, check email, and perform simple word processing tasks.
Would require help to install software or troubleshoot issues.
"Intermediate" users are comfortable with more complex tasks (e.g creating a spreadsheet or slide presentation).
Able to install/delete software programs and customize system settings without help.
"Advanced" users might have programming experience or someone others would consider a "techie".
Able to create a website or replace hardware components (e.g. hard drive) without help.
B2: What is your skill level as a typist?
Novice
Intermediate
Expert
"Novice" typists type slowly (< 20 words per minute) and must look at the keyboard to type.
"Intermediate" typists can type faster (25-40 words per minute) and do not have to look at the keyboard while typing.
"Expert" typists can type > 45 words per minute. Typing requires minimal effort and is preferred method of data entry.
Next


EHR User Survey - Section C

Tell us about your experience with your current EHR system...

C1: How long have you been using this EHR ? Enter number of years in the box. 5

If you have used the EHR < 1 year, enter a "M" after number of months in the box.
For your reference, the current EHR at the hospital was installed in 1996.

C2: How would you rate your expertise with this EHR? ○ Novice ○ Intermediate ◉ Advanced

"Novice" users are still learning the system and would not consider themselves proficient at this point.
Need to learn a great deal of skills and features that would make them much more productive users.

"Intermediate" users have become functional and reasonably proficient users of the system.
Still could learn some additional skills and features that might make them more productive users.

"Advanced" users have mastered most features/functions of the system and are highly proficient with the system.
Has realized all productivity gains from product. Would be considered a power user.

Next

Appendix E

Software evaluation survey

For *EHR Test Bench*

EHR Test Bench Software Evaluation

1. Do you find the videos of the clinical scenarios to be realistic and similar to real ED cases?

- ◯ Very realistic
- ◯ Moderately realistic
- ◯ Not very realistic

2. What do you think the ideal number of case videos that should be presented in a typical testing session (so not to tire out the test subject)?

- ◯ One case
- ◯ Two cases
- ◯ Three cases
- ◯ Four cases
- ◯ Five cases
- ◯ Six cases

3. The case scenarios viewed today showed relatively simple complaints and averaged three minutes in duration. Would you be willing to view more longer and more complicated clinical scenarios?

- ◯ Yes, I would like to see longer and more complicated case scenarios.
- ◯ No. These scenarios were complicated enough and I would not want to see longer videos.
- ◯ Not sure.

4. This initial testing session required an additional 20 minutes to orient the user and calibrate the software program. Future sessions with experienced testers would not require orientation and would be shorter. What do you think would be an ideal length of a testing session with experienced users?

- ◯ 15 minutes
- ◯ 30 minutes
- ◯ 45 minutes
- ◯ 60 minutes

Page 1

EHR Test Bench Software Evaluation

5. Did you have any problems understanding the onscreen instructions on how to complete the testing protocol?

- ◯ I understood the onscreen instructions without difficulty and did not have to ask questions of the exam proctor.
- ◯ I understood most of the onscreen instructions but had to ask some questions of the exam proctor.
- ◯ I did not understand the onscreen instructions and had to ask many questions of the exam proctor.

6. How user-friendly is the software's interface of EHR Test Bench?

- ◯ Very user-friendly
- ◯ Moderately user-friendly
- ◯ Slightly user-friendly
- ◯ Not at all user-friendly

7. Do you think this type of testing is an accurate method of testing electronic health record software?

- ◯ Yes.
- ◯ Maybe.
- ◯ No.

Other (please specify)

8. If this type of software testing were available over the internet and could be done from your home computer, would you consider participating in this type of testing?

- ◯ Yes
- ◯ Maybe
- ◯ No

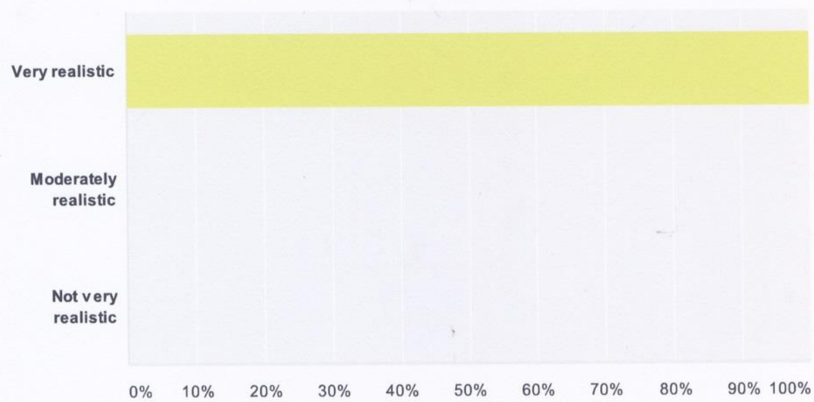
9. How can we improve our testing software?

10. Do you have any other comments or suggestions to our research team?

EHR Test Bench Software Evaluation

Q1 Do you find the videos of the clinical scenarios to be realistic and similar to real ED cases?

Answered: 10 Skipped: 0

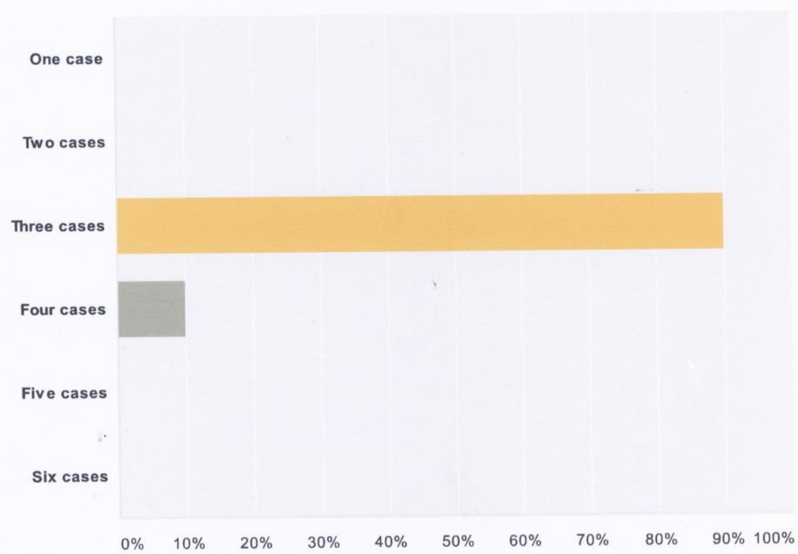


Answer Choices	Responses	
Very realistic	100.00%	10
Moderately realistic	0.00%	0
Not very realistic	0.00%	0
Total		10

EHR Test Bench Software Evaluation

Q2 What do you think the ideal number of case videos that should be presented in a typical testing session (so not to tire out the test subject)?

Answered: 10 Skipped: 0

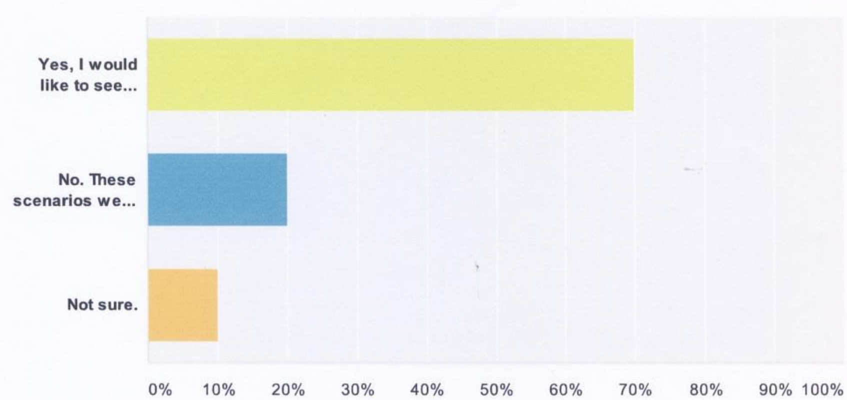


Answer Choices	Responses	
One case	0.00%	0
Two cases	0.00%	0
Three cases	90.00%	9
Four cases	10.00%	1
Five cases	0.00%	0
Six cases	0.00%	0
Total		**10**

EHR Test Bench Software Evaluation

Q3 The case scenarios viewed today showed relatively simple complaints and averaged three minutes in duration. Would you be willing to view more longer and more complicated clinical scenarios?

Answered: 10 Skipped: 0

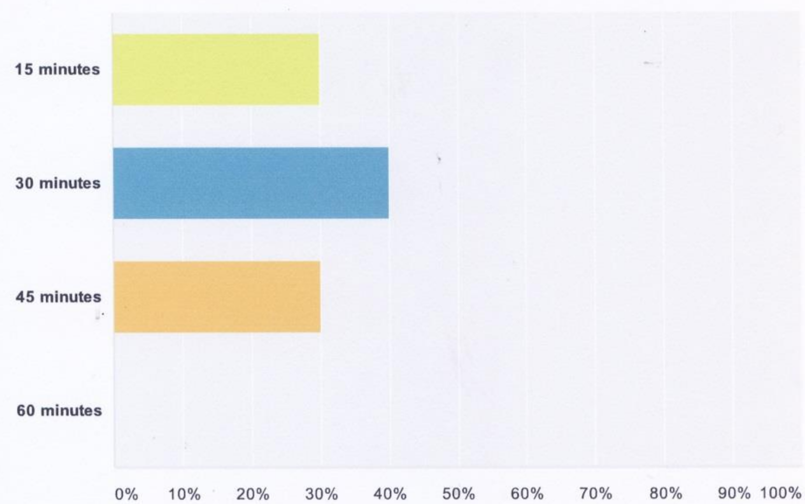


Answer Choices	Responses	
Yes, I would like to see longer and more complicated case scenarios.	70.00%	7
No. These scenarios were complicated enough and I would not want to see longer videos.	20.00%	2
Not sure.	10.00%	1
Total		10

EHR Test Bench Software Evaluation

Q4 This initial testing session required an additional 20 minutes to orient the user and calibrate the software program. Future sessions with experienced testers would not require orientation and would be shorter. What do you think would be an ideal length of a testing session with experienced users?

Answered: 10 Skipped: 0

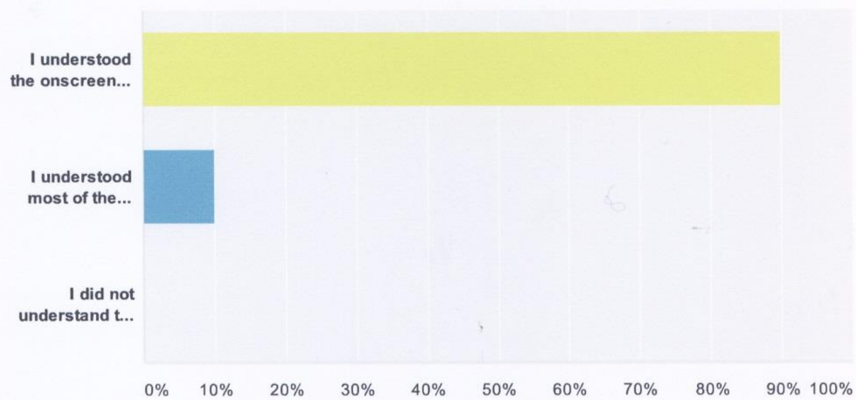


Answer Choices	Responses	
15 minutes	30.00%	3
30 minutes	40.00%	4
45 minutes	30.00%	3
60 minutes	0.00%	0
Total		**10**

EHR Test Bench Software Evaluation

Q5 Did you have any problems understanding the onscreen instructions on how to complete the testing protocol?

Answered: 10 Skipped: 0

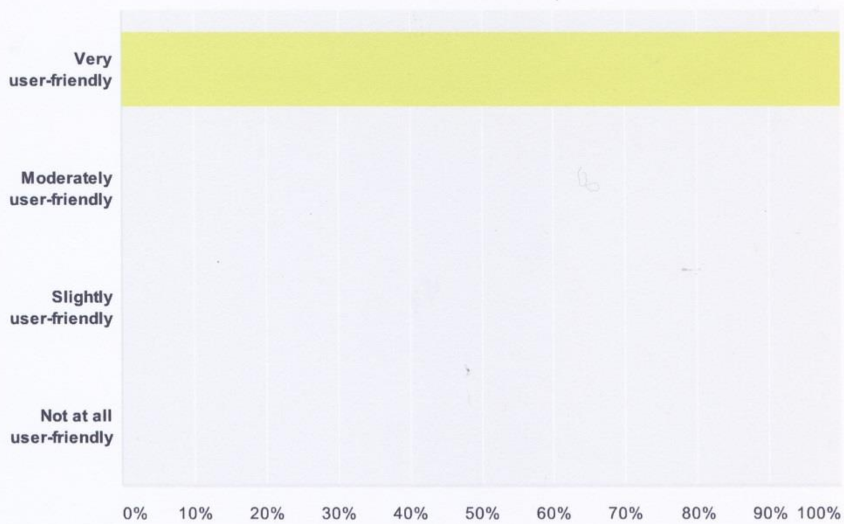


Answer Choices	Responses	
I understood the onscreen instructions without difficulty and did not have to ask questions of the exam proctor.	90.00%	9
I understood most of the onscreen instructions but had to ask some questions of the exam proctor.	10.00%	1
I did not understand the onscreen instructions and had to ask many questions of the exam proctor.	0.00%	0
Total		10

EHR Test Bench Software Evaluation

Q6 How user-friendly is the software's interface of EHR Test Bench?

Answered: 10 Skipped: 0

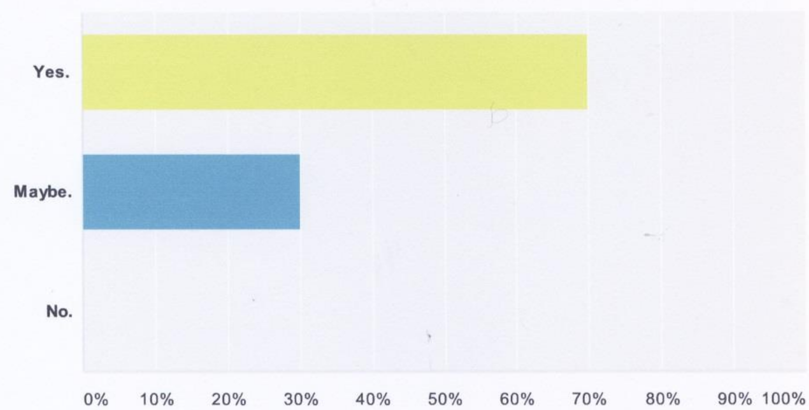


Answer Choices	Responses	
Very user-friendly	100.00%	10
Moderately user-friendly	0.00%	0
Slightly user-friendly	0.00%	0
Not at all user-friendly	0.00%	0
Total		**10**

EHR Test Bench Software Evaluation

Q7 Do you think this type of testing is an accurate method of testing electronic health record software?

Answered: 10 Skipped: 0

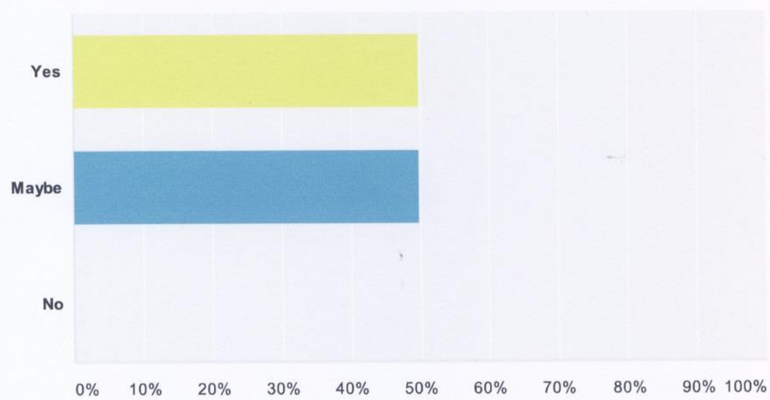


Answer Choices	Responses	
Yes.	70.00%	7
Maybe.	30.00%	3
No.	0.00%	0
Total		**10**

EHR Test Bench Software Evaluation

Q8 If this type of software testing were available over the internet and could be done from your home computer, would you consider participating in this type of testing?

Answered: 10 Skipped: 0



Answer Choices	Responses	
Yes	50.00%	5
Maybe	50.00%	5
No	0.00%	0
Total		10

References

[1] Friedman CP. J Am Med Inform Assoc. 2009; 16:169 –170. DOI 10.1197/jamia.M3092. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2649317/

[2] Lewis N. Healthcare IT pending to reach $40 billion. InformationWeek [Internet].May 16, 2011 [cited 2012 Sept 3]. Available from: http://www.informationweek.com/news/healthcare/EMR/229500682.

[3] Secrets of success on the EMR journey to meaningful use: leading hospital CIOs reveal key lessons learned.
Accenture [Internet]. 2011 Jan 4 [cited 2014 March 6]. Available at http://www.accenture.com/SiteCollectionDocuments/PDF/Secrets_to_Success_on_the_Journey_to_Meaningful_Use.pdf

[4] Classen DC, Bates DW. Finding the meaning in meaningful use. N Engl J Med 2011; 365:856. Available at http://www.nejm.org/doi/full/10.1056/NEJMsb1103659

[5] Kellerman AL, Jones SS. What it will take to achieve the as-yet-unfulfilled promises of health information technology. Health Affairs 2013; 32:65. Available at http://content.healthaffairs.org/content/32/1/63.full.html

[6] Sinsky CA, Hess J, Karsh B, Keller JP, Koppel R. Comparative user experiences of health IT products: how user experiences would be reported and used [Internet] Washington (DC): Institute of Medicine; 2012 Sep [cited 2014 Mar 6]. Available from: http://www.iom.edu/Global/Perspectives/2012/~/media/Files/Perspectives-Files/2012/Discussion-Papers/comparative-user-experiences.pdf

[7] Hill RG, Sears LM, Melanson SW. 4000 clicks: a productivity analysis of electronic medical records in a community hospital ED. Am J Emerg Med. 2013 Nov; 31(11):1591-4. doi: 10.1016/j.ajem.2013.06.028.

[8] Khorana AA. Physician as typist. J Clin Oncology. 2010; 28:3899-3900.

[9] Garcia-Castro R, Gomez-Perez A. Benchmarking in the semantic web. Semantic web engineering in the knowledge society. Cardoso J, Lytras M. editors. Hershey (PA): IGI Global; 2009. p 444.

[10] Hübner-Bloder G, Ammenwerth E. Key performance indicators to benchmark hospital information systems-a Delphi study. Methods Inf Med 2009; 48:508-518.

[11] Stevens A, Hernandez J, Johnsen K, Dickerson R, Raij A, Harrison C, DiPietro M, et.al. The use of virtual patients to teach medical students history taking and communication skills. Amer J Surg 2006; 191:806-811.

[12] Vendor Newsletter – HIMMS [Internet]. [cited 2012 May 23]. Available at: http://vendor.himss.org/HIMSS12/docs/HIMSS12_Summary.pdf

[13] Case L. AMD Radeon HD 7970 GHz Edition Review. PCWorld [Internet]. 2012 Jun 22 [cited 2014 Mar 7]; Available from: http://www.pcworld.com/article/258168/amd_radeon_hd_7970_ghz_edition_review.html

[14] Cerner: our emergency medicine solutions help emergency departments provide safe care efficiently [Internet].
2014 [cited 2014 Mar 6]. Available from:
https://www.cerner.com/solutions/Hospitals_and_Health_Systems/Emergency_Department/
[15]CCHIT.org [Internet]. [cited 2012 May 24]. Available at: https://www.cchit.org/about
[16] Federal Register Vol. 76, No. 5. pp.1262-1331. Friday, January 7, 2011 [cited 2014 Mar 6] Available at:
http://origin.www.gpo.gov/fdsys/pkg/FR-2011-01-07/pdf/2010-33174.pdf
[17]Gartner.com. Gartner methodologies [Internet]. [cited 2012 May 24]. Available at: http://imagesrv.gartner.com/research/methodologies/inside_gartner_research.pdf
[18]Klas Research the KLAS difference [Internet]. [cited 2014 Mar 6]. Available at: http://www.klasresearch.com/about/KLASDifference.aspx
[19] Klas Research KLAS data profile [Internet]. [cited 2014 Mar 6]. Available at: http://www.klasresearch.com/about/dataprofile.aspx
[20] Ammenwerth E, de Keizer N. An inventory of evaluation studies of information technology in healthcare trends in evaluation research 1982-2002. Methods Inf Med 2005; 44: 44–56 [cited 2014 Mar 5]. Available at: http://dare.uva.nl/document/47620
[21] Rahimi B, Vimarlund V. Methods to evaluate health information systems in healthcare settings: a literature review. J Med Syst 2007; 31:397–432 [cited 2014 Mar 7].
[22] Yen PY, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc. 2012 May-Jun; 19(3): 419 [cited 2014 Mar 7]. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3341772/
[23] Oxentenko AS, West CP, Popkave C, Weinberger SE, Kolar JC. Time spent on clinical documentation: a survey of internal medicine residents and program directors. Arch Intern Med 2010 Feb 22; 170(4):377-80. doi: 10.1001/archinternmed.2009.534.
[24] Ammenwerth E, Spotl HP. The time needed for clinical documentation versus direct patient care. Methods Inf Med. 2009; 48:84-91.
[25] Hollingsworth JC, Chisholm CD, Giles BK, Cordell W, Nelson DR. How do physicians and nurses spend their time in the emergency department? Ann Emerg Med 1998; 31:87-91.
[26] Hendrich A, Chow M, Skierczynski B, Lu Z. A 36-hospital time and motion study---how do medical-surgical nurses spend their time? Perm J 2008; 12:25-33.
[27] Poissant L, Pereira J, Tamblyn R, Kawasumi Y. The impact of electronic health records on time efficiency of physicians and nurses: a systematic review. J Am Med Inform Assoc. 2005; 12:505-16.
[28] Yee T, Needleman J, Pearson M, Parkerton P, Parkerton M, Wolstein J. The influence of integrated electronic medical records and computerized nursing notes on nurses' time spent in documentation. CIN 2012; 30:287-292.
[29] Read-Brown S, Sanders DS, Brown AS, Yackel TR, Choi D, Tu DC, Chiang MF. Time-motion analysis of clinical nursing documentation during implementation of an electronic operating room management system for ophthalmic surgery. AMIA Annu Symp Proc. 2013; 2013:1195-1204.
[30] Cheriff AD, Kapur AG, Qui M, Cole CL. Physician productivity and the ambulatory EHR in a large academic multi-specialty physician group. Int J Med Inform 2010; 79:492-500.

[31] Zhang J, Walji MF. TURF: toward a unified framework of EHR usability. J Biomed Inform 2011; 44:1056-1067.
[32] Turf – EHR usability toolkit. [Internet]. [cited 2014 Mar 11]. Available from: https://turf.shis.uth.tmc.edu/turfweb/
[33] Landman AB, Redden L, Neri P, Poole S, Horsky J, Raja AS, Pozner CN, Schiff G, Poon EG. Using a medical simulation center as an electronic health record usability laboratory. J Am Med Inform Assoc 2013; 0:1-6 Published online 2013 Nov 18. Doi:10.1136/amiajnl-2013-002233
[34] Zheng K, Padman R, Johnson MP, Diamond HS. An interface-driven analysis of user interactions with an electronic health records system. J Am Med Inform Assoc. 2009; 16:228-237.
[35] Hripcsak G, Vawdrey DK, Fred MR. Use of electronic clinical documentation: time spent and team interactions. J Am Med Inform Assoc. 2011; 18:112-117.
[36] Sanders DS, Lattin DJ, Read-Brown S, Tu DC, Wilson DJ, Hwang TS, Morrison JC, Yackel TR, Chiang MF. Electronic health record systems in ophthalmology---impact on clinical documentation. Ophthalmology 2013; 120:1745-1755.
[37] Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc 1997; 4:342-55.
[38] Logan JR, Gorman PN, Middleton B. Measuring the quality of medical records: a method for comparing completeness and correctness of clinical encounter data. Proc AMIA Symp. 2001; 2001:408-12.
[39] Chan P, Thyparampil PJ, Chiang MF. Accuracy and speed of electronic health record versus paper-based ophthalmic documentation strategies. Am J Ophthalmol 2013; 156:165-172.
[40] Hammond KW, Helbig ST, Benson CC, Brathwaite-Sketoe BM. Are electronic health records trustworthy? Observations on copying, pasting and duplication. AMIA Annu Symp Proc. 2003; 2003: 269–273.
[41] Thornton JD, Schold JD, Venkateshaiah L, Lander B. Prevalence of copied information by attendings and residents in critical care progress notes. Crit Care Med 2013; 41(2):382-8.
[42] Friedman CP, Wyatt JC. Evaluation methods in biomedical informatics. 2nd edition. New York: Springer; 2010. Chapter 5, Measurement fundamentals; p.126.