

Usability characteristics of self-administered computer-assisted interviewing in the emergency department

Factors affecting ease of use, efficiency, and entry error

D. B. Herrick^{*1}; A. Nakhasi^{*2}; B. Nelson³; S. Rice¹; P. A. Abbott⁴; A. S. Saber Tehrani¹; R. E. Rothman⁵; H. P. Lehmann⁶; D. E. Newman-Toker¹

¹ Johns Hopkins University School of Medicine, Department of Neurology; ² Johns Hopkins University School of Medicine; ³ Johns Hopkins University; ⁴ Johns Hopkins University School of Nursing; ⁵ Johns Hopkins University School of Medicine, Department of Emergency Medicine; ⁶ Johns Hopkins University School of Medicine, Division of Health Sciences Informatics

*These authors request co-first authorship

Keywords

Medical history taking, computer assisted diagnosis, clinical decision-support systems, point of care systems, triage

Summary

Objective: Self-administered computer-assisted interviewing (SACAI) gathers accurate information from patients and could facilitate Emergency Department (ED) diagnosis. As part of an ongoing research effort whose long-range goal is to develop automated medical interviewing for diagnostic decision support, we explored usability attributes of SACAI in the ED.

Methods: Cross-sectional study at two urban, academic EDs. Convenience sample recruited daily over six weeks. Adult, non-level I trauma patients were eligible. We collected data on ease of use (self-reported difficulty, researcher documented need for help), efficiency (mean time-per-click on a standardized interview segment), and error (self-report age mismatched with age derived from electronic health records) when using SACAI on three different instruments: Elo TouchSystems ESY15A2 (finger touch), Toshiba M200 (with digitizer pen), and Motion C5 (with digitizer pen). We calculated descriptive statistics and used regression analysis to evaluate the impact of patient and computer factors on time-per-click.

Results: 841 participants completed all SACAI questions. Few (<1%) thought using the touch computer to ascertain medical information was difficult. Most (86%) required no assistance. Participants needing help were older (54 ± 19 vs. 40 ± 15 years, $p < 0.001$) and more often lacked internet at home (13.4% vs. 7.3%, $p = 0.004$). On multivariate analysis, female sex ($p < 0.001$), White ($p < 0.001$) and other ($p = 0.05$) race (vs. Black race), younger age ($p < 0.001$), internet access at home ($p < 0.001$), high school graduation ($p = 0.04$), and touch screen entry (vs. digitizer pen) ($p = 0.01$) were independent predictors of decreased time-per-click. Participant misclick errors were infrequent, but, in our sample, occurred only during interviews using a digitizer pen rather than a finger touch-screen interface (1.9% vs. 0%, $p = 0.09$).

Discussion: Our results support the facility of interactions between ED patients and SACAI. Demographic factors associated with need for assistance or slower interviews could serve as important triggers to offering human support for SACAI interviews during implementation.

Conclusion: Understanding human-computer interactions in real-world clinical settings is essential to implementing automated interviewing as means to a larger long-term goal of enhancing clinical care, diagnostic accuracy, and patient safety.

Correspondence to:

David E. Newman-Toker
Johns Hopkins Hospital
Meyer 8-154
600 North Wolfe Street
Baltimore, MD 21287
email: toker@jhu.edu
tel: 410-502-6270
fax: 410-502-6265

Appl Clin Inform 2013; 4: 276–292

DOI: 10.4338/ACI-2012-09-RA-0034

received: September 12, 2012

accepted: June 1, 2013

published: June 19, 2013

Citation: Herrick DB, Nakhasi A, Nelson B, Rice S, Abbott PA, Tehrani ASS, Rothman RE, Lehmann HP, Newman-Toker DE. Usability characteristics of self-administered computer-assisted interviewing in the emergency department. Factors affecting ease of use, efficiency, and entry error. *Appl Clin Inf* 2013; 4: 276–292
<http://dx.doi.org/10.4338/ACI-2012-09-RA-0034>

1. Introduction

Misdiagnosis is a major public health problem and an area in need of more research towards systems-oriented solutions [1]. Nearly half of all major diagnostic errors result from failures of clinician assessment, among these faulty medical history-taking (~10%), bedside examination (~10%), and bedside decision logic (~30%) [2]. In the emergency department (ED), misdiagnosis appears to account for the majority of errors [3, 4], with many involving serious injury or death [5].

Diagnostic decision support has been suggested as a powerful tool to combat misdiagnosis and improve patient safety [6]. General decision support systems have been rigorously developed in computer science laboratories [6], but are not used largely as a result of inattention to workflow. Although there has been little formal study of success factors specific to diagnostic decision support systems, a recent systematic review found that the feature most predictive of any decision support's success was "automatic provision of decision support as part of clinician workflow" [7].

An approach that may offer a greater chance for success in the integration of diagnostic decision support into the ED workflow is to gather data directly from patients prior to their medical encounter, without a physician present [8]. Self-administered computer-assisted interviewing (SACAI) has already demonstrated accuracy and little difficulty [9] in obtaining clinical data [10, 11, 12], so could potentially reduce diagnostic error by reducing faulty history-taking. Nevertheless, there has been relatively little study of SACAI in the ED setting [8, 13-20] and almost none on its use for the specific purpose of diagnosis [8, 19, 20]. In an effort to establish a workflow-sensitive approach to diagnostic decision support in the ED, we have sought to study complaint-specific, adaptive interviews to seek key elements of the medical history from patients in a pre-encounter setting (e.g., ED waiting area).

As part of an ongoing research with the long-term goal of developing automated medical interviewing for diagnostic decision support, we first determined the willingness of participants to engage in SACAI, that will be reported in a related manuscript (Nahkasi et al., unpublished data, in submission). In this analysis, we sought to elucidate the usability dynamics between participants and SACAI by analyzing indicators for ease, error, and efficiency. Usability can be defined as the ease with which a system allows its user to achieve their goals [21, 22]. More specifically, we explored whether participants were able to engage in SACAI with little difficulty, act independently, pay attention, honestly and accurately answer questions, demonstrate low rates of misclick errors, and utilize SACAI in a time-efficient manner.

2. Methods

This report is part of an ongoing research effort whose long-range goal is to develop automated medical interviewing for diagnostic decision support (AHRQ HS017755-01). The methods and procedures of this study are described in greater detail in the related report (Nahkasi et al., unpublished data, in submission). Briefly, we conducted a cross-sectional study concurrently at 2 urban academic EDs. A convenience sample was recruited daily from 9 a.m. to 9 p.m., 7 days per week, during a six-week period (July-August, 2009). All adult, non-level I trauma patients were eligible and exclusions were principally for altered mental status, risk of violence, and illness severity. We administered brief (~6 minute, ~40 items), adaptive, SACAI interviews regarding medical symptoms related to the ED visit. Subjects used one of three interactive technologies: Elo TouchSystems ESY15A2 (15" Thin Film Transistor (TFT) Active Matrix Panel, AccuTouch finger touch technology, all-in-one personal computer set up as a roving kiosk using a mobile computer workstation); Toshiba Portege M200 (4.4 lbs. (2.0 kg) convertible laptop-style tablet personal computer with 11.6" display, used with a tablet digitizer pen); Motion Computing C5 (3.3 lbs. (1.5 kg) clipboard-type tablet personal computer with 10.4" XGA TFT Advanced Fringe Field Switching Technology LED Backlight Display, used with a digitizer pen) [22]. Assignment to one of the two hardware technologies was non-random, and based on location of interview (digitizer pen interviews were generally conducted in individual patient rooms while the touch screen interview using a roving kiosk was conducted in the main ED waiting area). Adaptive interviews were designed and delivered using a commercially available software package that allows for multiple question types and formats for presentation (Digivey Survey

Suite, version 3.1; Creoso, Phoenix, AZ). This software was originally designed for use in conducting brief shopping mall kiosk surveys, but has been adapted for use in clinical research. Patients seated in the ED waiting area were approached with an adjustable-height computer workstation or hand-held tablet computer, while those in hospital beds had the portable computer placed on a rolling food tray. The primary SACAI interview was conducted under the supervision of one of 12 trained, college-level research assistants who were instructed not to assist patients after introducing them to the study and interview process. Patients were offered the option of answering 'practice' questions before officially starting the medical interview. For all patients, the interview consisted of two discrete blocks addressing the same interview questions: a human interview (researcher-administered questions using the portable hardware/software as a data entry tool) and SACAI (patients interacting directly with the hardware/software without interference by the researcher). The order of interview blocks (SACAI versus human first) was randomly assigned. The randomization was weighted towards having the SACAI interview precede the human interview in a 3:1 ratio, since our primary goal was to assess how SACAI performed when conducted in a "pre-encounter setting (i.e. before a patient was asked medical history questions by a health care professional). All question variations described below were prospectively planned and included random allocation of participants to each subgroup. All study data were gathered electronically.

2.1. Study Procedures

To assess the interactions between participants and SACAI, participants were asked multiple usability-related questions throughout the interview, some administered via SACAI, others by human interview. Research assistants were also asked to evaluate these interactions through direct observation. To capture the level of difficulty of SACAI, participants were asked in the self-administered portion of the interview: "How hard was it for you to use this touch computer?" Response options included "Very Easy", "Somewhat Easy", "In Between Easy and Hard", "Somewhat Hard", "Very Hard". The research assistant asked the participant, "Were you able to pay attention the whole time you were answering questions on the computer?" Response options included "Yes; completely", "Yes; mostly", "No". To assess the level of participant independence (an external, non-self-report measure of ease of use) while taking the survey, the research assistant was asked at the conclusion of the interview: "Did the patient require assistance to complete the self-administered part of the interview?" Response options were "Enormous," "Moderate," "Minimal," and "None". Follow-up questions detailed who helped the participant, why the participant required assistance, and other clinical variables.

To assess how honestly and accurately participants believed they answered questions during SACAI, participants were asked two questions: 1) "Did you answer all the questions truthfully and accurately to the best of your ability?" and 2) "Did you make any mistakes when you were answering questions on the computer?" Participants were also questioned by research assistants as to whether they had misclicked a response during SACAI. The purpose behind asking these questions was to determine whether self-reported errors or dishonesty could be used someday as a "filter" to invalidate SACAI-generated responses that might be error-prone. We independently verified the accuracy of patient responses (i.e., misclicks) to a self-report question about age. During the SACAI interview, participants were asked to enter their age and race. Age was entered using a numeric keypad displayed on the screen. We randomized participants to one of three age entry fields, each having different input restrictions, to test whether fields with more stringent data validation rules collected more accurate data. The age fields had one of the following restrictions: (a) no limit on number of digits or numeric value, (b) 3-digit limit and numeric value of 1-120, or (c) 2-digit limit and numeric value of 15-99. Research assistants recorded the patient's name, date of birth, and Medical Record Number (when available). Error rates were calculated by comparing self-reported age entry through SACAI against hospital records, whenever available. An age entered was considered incorrect if it was more than one year different than the participant's current age as determined from the hospital record.

To investigate the impact of question and answer choice format on response consistency, three target 'review of symptoms' items were prospectively chosen ("stuffy or runny nose", "sore throat," and "ear ache"). The SACAI interview questioned participants as to whether they had experienced

these specific medical symptoms over the last seven days, but presented the question(s) and answers in one of five possible formats. Participants were randomly assigned to question presentation with these three symptoms either (a) one-at-a-time, in sequence, in a binary-response format (i.e., “yes” or “no”; 1 screen per symptom inquiry); or (b) together in a multiple-response format (i.e., choose the symptoms you have experienced) that included a randomly varying number of ‘distractor’ symptoms presented on the same screen alongside the three target symptoms (total symptoms per screen 4, 6, 9, or 12). Distractor response options included “sore throat,” “wheezing,” burning with urination,” “fever,” “chills or shakes,” “achey joints or muscles ‘all over,’” “unusual sweating,” “rash or hives,” or “None of these.”

We also sought to analyze which factors influenced how long it took participants to complete SACAI. For comparison across all participants, various time points were recorded (as computer-generated time stamps) and used to determine the amount of time it took to complete a uniform portion of the interview.

Basic demographic data were gathered, including sex and age (which could be externally verified from the electronic medical record). Race and ethnicity were recorded using standard National Institutes of Health (NIH) guidelines; when subjects identified themselves with more than one race, they were prompted to choose their preferred racial identity.

At the conclusion of the primary data collection phase, 7 of the 12 research assistants were interviewed about their subjective experiences interacting with patients and the SACAI system. Interviews were designed in five categories: reasons for declining, form factor, input, time to completion, patient assessment, and ease of use. Data were analyzed qualitatively. The reasons for declining are discussed in the related report on willingness to participate (Nakhasi et al., unpublished data, in submission).

2.2. Data Analysis

We calculated descriptive statistics regarding level of difficulty, independence, honesty, accuracy, and influence of question and answer format on recording of medical symptoms. Participant responses to the question “How hard was it for you to use this touch computer?” were mapped to a binary variable (difficult, not difficult) with “Somewhat Hard” and “Very Hard” categorized as difficult. We used a χ^2 test for trend across ordered groups to determine whether question and answer choice format (i.e., number of response options presented – binary, 2, 4, 6, 9, 12) influenced how participants responded regarding their three target review of symptoms questions.

From computer-generated time stamps, we measured the total time to complete a specific, standardized segment of the SACAI interview. This segment of the interview had no adaptive logic branches, so participants answered the same total number of questions (one question per screen). However, there were several questions in a multiple-response format (e.g., “select all symptoms that apply”), so the total number of ‘responses’ varied across subjects. We therefore calculated the average time-per-click (i.e., per response) to account for disparities in total time based solely on having chosen more response options. Typical response options were “yes-no,” Likert-type ordinal scales, or involved choosing from a list. This measure was stored as a continuous variable (time-per-click). Data from participants who had interrupted interviews (defined as a survey paused manually by the research assistant, flagged automatically by the survey software) were not included in this time-per-click analysis.

Balance across randomized groups (human-first vs. SACAI-first) was assessed to ensure comparability. Secondary analyses were conducted to test for significant demographic (sex, race, age, education) or other (elective practice questions, current pain level) heterogeneity across randomized groups. The Kruskal-Wallis rank-test was used when comparing differences in age across randomized groups. Categorical variables (sex, race, internet access, and education) were compared across randomized groups using the χ^2 statistic.

NIH race categories were mapped to three mutually exclusive subgroups (black, white, other); using the participant’s preferred racial identity if more than one was selected. Participant ages were mapped to three groups for analysis: 18-40 years, 41-65 years, and >65 years. Pain level, 0 being the lowest and 10 being the highest, was mapped to three groups for analysis: 0-3, 4-6, and 7-10. Univariate and multivariate regression analyses were used to evaluate how time-per-click was in-

fluenced by participant sex, race, age, level of pain, mode of answer selection (touch-screen or digitizer pen), randomization to human vs. SACAI first, internet access, and level of education. Logistic regression analysis was used to evaluate how these same factors influenced ease of use and error.

Data were exported from Digivey Survey Suite into Microsoft Excel 2007 (Microsoft, Redmond, WA). For coding and statistical analyses, data were transferred into STATA statistical software, version 11 (STATA, College Station, TX) using Stat Transfer Version 10 (Circle Systems, Seattle, WA). All p-values were 2-sided, with $p < 0.05$ considered statistically significant. All reported n values denote the number of subjects.

3. Results

Of the 1,705 patients approached for screening, 1,027 were recruited and 841 completed the self-administered portion of the survey. Interviewees could not “skip” questions, so the remaining subjects stopped before completing the interview, usually because patient care intervened; results from partial interviews are not reported here. Demographic characteristics of patients are reported in ► Table 1. SACAI was administered by patients using a tablet personal computer with digitizer pen ($n = 648$, mostly in the ED waiting area, ED patient rooms in one of the non-level I bays, or Emergency Acute Care Unit) or touch-screen monitor ($n = 193$, almost all in the ED waiting area). Only 574 patient ages could be independently verified using the electronic medical record (in the other cases, medical record numbers were unavailable).

When asked how hard it was to use the touch computer, 92% ($n = 776/841$) of participants reported “Very Easy,” 5% ($n = 45/841$) reported “Somewhat Easy,” and 1% ($n = 12/841$) reported “Between Easy/Hard,” while fewer than 1% ($n = 8/841$) answered “Somewhat Hard” or “Very Hard” (► Figure 1). Only 11% ($n = 91/841$) of participants chose to take practice questions prior to starting the self-administered portion of the interview, and there was no correlation between choosing to take practice questions and the self-reported level of difficulty. However, of the participants who thought the touch computer was difficult to use, fewer (20%, $n = 4/20$) reported having internet at home compared to participants who thought the touch computer was “Between Easy/Hard” or easier (56%, $n = 456/821$, $p = 0.002$). Participants who thought the interview was difficult were also less efficient based on time-per-click (11.7 ± 6.1 vs 8.4 ± 3.6 seconds, $p < 0.001$) and were less likely to use the touch computer completely unassisted (70% unassisted, $n = 14/20$ vs. 90% unassisted, $n = 710/785$, $p < 0.001$) as illustrated in ► Figure 1. Demographic, clinical (pain), and equipment-based (digitizer pen vs. touch screen) variation in ease of use, error rate, and efficiency (time-per-click) is shown in ► Table 2. Multivariate analysis shows female sex ($p < 0.001$), White ($p < 0.001$) and other ($p = 0.05$) race (vs. Black race), younger age ($p < 0.001$), internet access at home ($p < 0.001$), high school graduation ($p = 0.04$), and touch screen entry (vs. digitizer pen) ($p = 0.01$) are independent predictors of decreased time-per-click. There was no statistically significant difference in time-per-click between the different computers that used digitizer pens. Pain level was not associated with time-per-click in this analysis.

As reported by the research assistant on completion of the encounter, 90% ($n = 724/805$) of participants did not require assistance to complete the SACAI portion of the interview. Those requiring assistance were considered to need “Minimal” (5%, $n = 43/805$), “Moderate” (2%, $n = 15/805$), or “Enormous” (3%, $n = 23/805$) assistance. Research assistant responses to this question were unavailable for 4% (36/841) of participants due to early termination. ► Figure 1 illustrates that a larger fraction of participants requiring moderate or enormous help reported the touch computer to be “Very Hard” (8%, $n = 3/38$ vs. 1%, $n = 4/767$) and a smaller proportion reported the touch screen to be “Easy” to use (76%, $n = 29/38$ vs. 93%, $n = 712/767$) compared to those who did not require assistance ($p < 0.001$). ► Figure 2 illustrates the major reasons for patients’ need for assistance as reported by research assistants. Participants who needed help from others were older (54 ± 19 vs. 40 ± 15 years, $p < 0.001$) and a larger proportion did not have internet at home (13.4% vs. 7.3%, $p = 0.004$). Participants who used the touch-screen required less help than those who used the digitizer pen ($p = 0.023$); although not statistically significant, similar trends towards better touch-screen usability were seen in the ease, error, and efficiency analyses (► Table 2).

Only 3% (5/158) of participants reported they could not pay attention the entire time they were answering questions on the computer. When asked if all questions were answered truthfully and accurately to the best of their ability, 92% (n = 775/841) of participants reported "Yes; best I could," 7% (n = 59/841) answered "Yes; mostly," and 1% (n = 7/841) answered "No." When questioned through SACAI whether any mistakes were made in answering questions on the computer, 91% (769/841) reported "No," while 8% (71/841) answered "Yes; a few mistakes," and one participant (1/841) reported "Yes; a lot of mistakes." When asked by the research assistant, 0.2% of participants indicated that they had mislicked a response. The calculated mislick rate based on self-reported age entry using an on-screen keypad compared against hospital records was 1.92% (n = 11/574). Mislicks were not more frequent among those who self-reported having made a few mistakes (2.4% vs. 1.9%, p = 0.95). Age entry mislick error was not associated with ease of use based on self-reported level of difficulty. There was no statistically significant association between age-entry mislick and randomly-assigned field data entry restrictions. The mislick rates for the unrestricted, 2-digit limit, and 3-digit limit conditions were 1.6% (n = 3/191), 1.7% (n = 3/181), and 2.5% (n = 5/202), respectively. No age entry mislicks were made among participants who used the touch-screen monitor, compared to 11 mistakes made among participants who used the tablet computer with digitizer pen (χ^2 , p = 0.09). There was no statistically significant difference in mislick rates between the different computers that used digitizer pens.

Participant responses regarding presence or absence of the three targeted symptoms differed between the binary (n = 168) and multiple-response choice (n = 673) formats (► Figure 3). All three symptoms were reported more frequently when inquired about individually than when selected from a list of options "stuffy or runny nose" (29% vs. 21%, respectively, p = 0.032), "sore throat" (17% vs. 9%, p = 0.005), and "ear ache" (13% vs. 9%, p = 0.058). Additionally, as the number of distractor symptoms increased within the multiple-response format (1 to 3 to 6 to 9 distractors), the "sore throat" symptom demonstrated a decreased likelihood of being chosen (p = 0.015, χ^2 test for trend); "stuffy or runny nose" and "ear ache" demonstrated a similar trend (► Figure 3), but neither was statistically significant.

The mean time for participants to complete the self-administered portion of the interview was 5.78 minutes (n = 814; range 1.15 – 27.38; median 5.22; interquartile range 4.15 – 6.85) and the mean number of questions answered was 40. Twenty-seven interviews were terminated early prior to the final time point collection and were not included in this analysis. The participant who took 27.38 minutes to complete the SACAI interview was a 61 year-old Caucasian male who reported visiting the ED due to chest pain, trouble breathing, belly pain, neck pain, back pain, and "other reasons," self-reported as "deep leg thrombosis." The patient reported a pain level of 9 on a 1-10 scale, and did not report SACAI as being difficult when asked on the computer.

Multivariate regression analyses demonstrated that efficiency (time-per-click) was associated with sex, race, age, internet access at home, and high school education (► Table 2). The associations in ► Table 2 were still present in bivariate and multivariate models taking into account other potential confounders (e.g., triage severity level, chief complaint). There was no association between efficiency and whether the participant was randomized to SACAI before human interview or the reverse, the participant's level of pain, or whether the participant used the touch-screen monitor or digitizer pen. Time-per-click was not associated with error based on externally-verified age entry mislick rates.

Following the self-administered portion of the interview, 86% (n = 726/841) of participants indicated that SACAI was either a "Fantastic" or "Good" idea to help get medical information from patients.

In the qualitative analysis, seven research assistants responded to questions about usability. When asked for feedback on overall usability of the SACAI system, all expressed concerns about bulkiness and mobility difficulties with the roving kiosk. Nevertheless, 2 of 7 believed this setup improved ease of use for patients in areas that could accommodate the size. The touch-screen monitor was described by 6 of 7 research assistants as simple for the participants to use; 2 of 7 noted that the digitizer pen decreased ease of use. Concerns regarding privacy were raised by two and concerns regarding sanitization methods were raised by one. Increase in required strength and dexterity to support and operate the tablet and digitizer pen (n = 3), patients' fear of dropping the tablet (n = 1) and the

tablet's decreased usability in reclined patients ($n = 1$) were among other issues brought up by the research assistants.

4. Discussion

From our study, SACAI in the ED produced highly encouraging usability ratings across the domains of ease of use, efficiency, and error based on misclick rates. Our results generally accord with prior studies that have found a high degree of acceptability of SACAI in the ED setting [8, 13], and we describe several usability attributes not previously reported. Participants using SACAI found this technology "Very Easy" to use (92%), misclicked at a low rate (1.92%) even when using an on-screen numerical keypad without data validation rules, and, on average, completed 40 questions during the interview in under 6 minutes. The overwhelming majority (86%) thought that SACAI was either a "Fantastic" or "Good" idea to help obtain medical information. We identified five other usability-related findings from this study that may have design implications for future systems.

First, special design features may be needed to effectively engage those with physical disabilities. Ease of use for any developed SACAI system should not simply be tailored for the average participant. Instead the system's usability should address the full range of potential users in the relevant clinical setting, in this case the ED. Our system appeared to largely satisfy this variation in the user population we tested, with 86% of participants requiring no assistance and only 9% of participants requiring "Moderate" or "Enormous" help. This result, however, should be interpreted cautiously given that only about 50% of eligible ED patients participated and completed the interview. It is encouraging that those who chose not to take any practice questions beforehand (89%) reported the same degree of difficulty in using SACAI as those who took practice questions. Nevertheless, it is important to note that patients did not effectively identify their own needs for assistance simply by answering the self-administered question about desire for practice or the one about ease of use of the SACAI system – more than half of those requiring at least moderate assistance rated the SACAI interview "Easy". Of those requiring at least moderate assistance, 76% were deemed to have vision, cognitive, or motor problems as their reasons for needing help. To reduce these burdens, potential areas for design improvements include greater readability through font size, color, or style customization; computer vocalization of questions and answer choices; availability of question-specific instructions; adjustment of touch screen sensitivity; and incorporation of oral response input to reduce the need for motor exertion. Further investigation regarding the influence of these usability-related design and feature modifications is warranted.

Second, certain usability design characteristics we tested had more influence on response accuracy than others. While data field restrictions may be helpful in certain contexts [23, 24], they did not result in any change to the rate of age entry mistakes in our study. Interestingly, in this study all age entry mistakes ($n = 11$) were made using the digitizer pen technology and none were made with the touchscreen interface. While this may represent a chance or biased finding in this small sample of errors obtained from these two non-randomly allocated subgroups, qualitative descriptions from the research assistants appear to corroborate that the pen technology was more difficult for patients to use. Further study is warranted to determine if the digitizer pen truly lowers the accuracy of the data entry process. If so, the use of a touch-screen monitor could minimize error when electronically collecting patient data in the ED. While this hardware feature may have influenced click accuracy slightly, a question and answer design feature we tested definitely biased patient responses with regard to review of symptoms reporting. The format of question and answer choices (binary versus multiple-response format) resulted in inconsistent symptom prevalence estimates, with the binary format (one question per symptom) garnering higher affirmative responses. We have confidence in these findings because patients were randomly assigned to these subgroups in sufficient numbers. This may be an important consideration when assessing medical symptoms using SACAI, and further research is needed in order to discern which format produces the more accurate and more medically-relevant responses.

Third, our SACAI system offers promise that data entry speeds are probably realistic for ED use. The average time spent answering SACAI questions in this study was below 6 minutes for a mean of 40 questions. This value is well below patients' median reported 'willingness to participate' threshold

of 60 minutes (Nakhasi et al., unpublished data, in submission) and the average ED wait time of 50 minutes nationally [25]. Extrapolating from our present results to a 30+ minute SACAI interview, it might be possible to ask patients an average of 200 questions or more, which could be sufficient to assist in diagnosis through a pre-encounter, automated, symptom-specific interview strategy. Using a simple, interactive visual design (one that is used in mall kiosks around the country) may have contributed to the understandability of SACAI, a key component of usability [26]. Given that patients would prefer to spend the ED waiting time answering relevant medical questions over simply waiting (Nakhasi et al., unpublished data, in submission), it is conceivable that converting ED waiting room time to SACAI interviews might even improve overall patient satisfaction in the ED setting.

Fourth, in looking across all demographic characteristics associated with participant usability, access to the internet proved to be the most critical. Those with access demonstrated greater ease of use, lower error rates based on age-entry misclicks, and improved efficiency. This could be attributed to this population's greater familiarity with computers and data entry methods. Access and familiarity may also be correlated with age and choice to participate, given the slightly reduced fraction of patients over age 65 in our population (9.3%) relative to a representative sample of ED patients (14.5%) [27]. Promisingly, over the last five years, the percentage of adults 65 years and over using the internet has increased more than 55% from 11 to 17 million [28] and continues to grow. Additionally, public schools with internet access increased from 3% in 1994 to 94% in 2005 [29], which will provide future adults with more computer and internet experience. However, at the moment, not having access could serve as an important trigger to identify the need for additional assistance while engaging SACAI. The largest differences in efficiency (seconds per click) were seen across age groupings, with those in the over 65 group showing a 42% longer mean per-click time relative to those under age 40. This may suggest that older individuals may require more time than younger individuals to complete fixed-length SACAI interviews. There were smaller but still statistically significant differences in SACAI efficiency by sex, race, and high school education that persisted in multivariate models controlling for known potential confounders. Whether these demographic differences are real or due to unmeasured possible confounders remains unclear and could be an area important for future study.

Finally, reducing the size of the SACAI system to maximize portability, using privacy screen filters that narrow the viewing angle so that the data are only visible directly in front of the monitor, and putting specific sanitization protocols in place are among ways to address usability concerns raised by research assistants. These concerns may be as important to healthcare providers (e.g., nurses, hospital administrators, infection-control workers) as to patients.

We identified several limitations of our study. There was non-random and unbalanced assignment of participants to touchscreens versus digital pens. Our results may better reflect usability characteristics of SACAI using digitizer pen hardware rather than touch screen technology, since roughly three of every four participants used the pen system. Verification of user misclicks based on age could only be completed in 68% (574/841), and other responses (e.g., "true" symptom profiles) could not be verified. Participants could have inflated their reports of ease of use based on social desirability (30), given dishonest responses in an attempt to decrease their wait times, or clicked through the results quickly because they did not care about their responses. Our results might not be generalizable to all EDs or all ED patients, since only 60% ($n = 1,207/1,705$) at our two EDs participated. Non-participants probably differed from participants, and our data may not generalize well to non-participants.

5. Conclusion

By furthering our understanding the nature of human-computer interactions in real-world clinical settings, we move one step closer to validating automated interviewing as means to our ongoing long-term goal of enhancing clinical care, diagnostic accuracy, and ultimately patient safety. In this study, we were able to show that participants were able to engage with SACAI as independent, efficient actors with self-reported and observer-determined high levels of ease and apparently low levels of user input error. The usability characteristics described in this report bolster the willingness to

participate data presented in the related report (Nakhasi et al., unpublished data, in submission). Together, these reports suggest that kiosk-based, pre-encounter automated medical interviewing may be a realistic, viable approach to pre-encounter diagnostic decision support in pursuit of a workflow-sensitive solution to ED misdiagnosis. Future studies should seek to optimize usability for those with physical limitations in function and tailor question design to maximize accuracy and medical relevance of patient responses. We plan next to examine the reliability and validity of automated interviewing relative to human interview in the interpretation of patient symptoms.

Implications of Results

We have shown that patients in the emergency department setting are able to engage with SACAI independently and efficiently with what appear to be low levels of error. This suggests that it may be feasible to utilize kiosk-based, pre-encounter automated medical interviewing for pre-encounter diagnostic decision support as a step toward a workflow-sensitive solution to ED misdiagnosis.

Responsibility for Manuscript

The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Conflict of Interest Statement

No conflicts of interest. None of the authors have any financial or personal relationships with other people or organizations that could inappropriately influence (bias) their work.

Role of Medical Writer or Editor

No medical writer or editor was involved in the creation of this manuscript.

IRB Approval

Ethics approval was provided by the Johns Hopkins Hospital Institutional Review Board.

Sources of funding and support; an explanation of the role of sponsor(s):

This study was supported principally by a federal grant from the Agency for Healthcare Research and Quality (HS017755-01).

Information on previous presentation of the information reported in the manuscript

No information reported in this manuscript was previously presented.

All persons who have made substantial contributions to the work but who are not authors

Acknowledgments: We would like to extend our deepest gratitude to the dedicated research assistants whose hard work is reflected in the results of this study: Ashlin Burkholder, Mehdi M. Draoua, Chinedu Jon-Emefieh, Lauren Hawkins, Bernhard A. Kurth, Karlee Lau, Yali Lin, Benjamin Nelson, John Prendergass, Liandra N. Presser. We would also like to thank Jakob and Elisabeth Scherer of Creoso, Inc. for their generous assistance in optimizing their software for use in a clinical research setting, without which this study would not have been possible.

Written permission obtained by corresponding author for those who are being acknowledged.

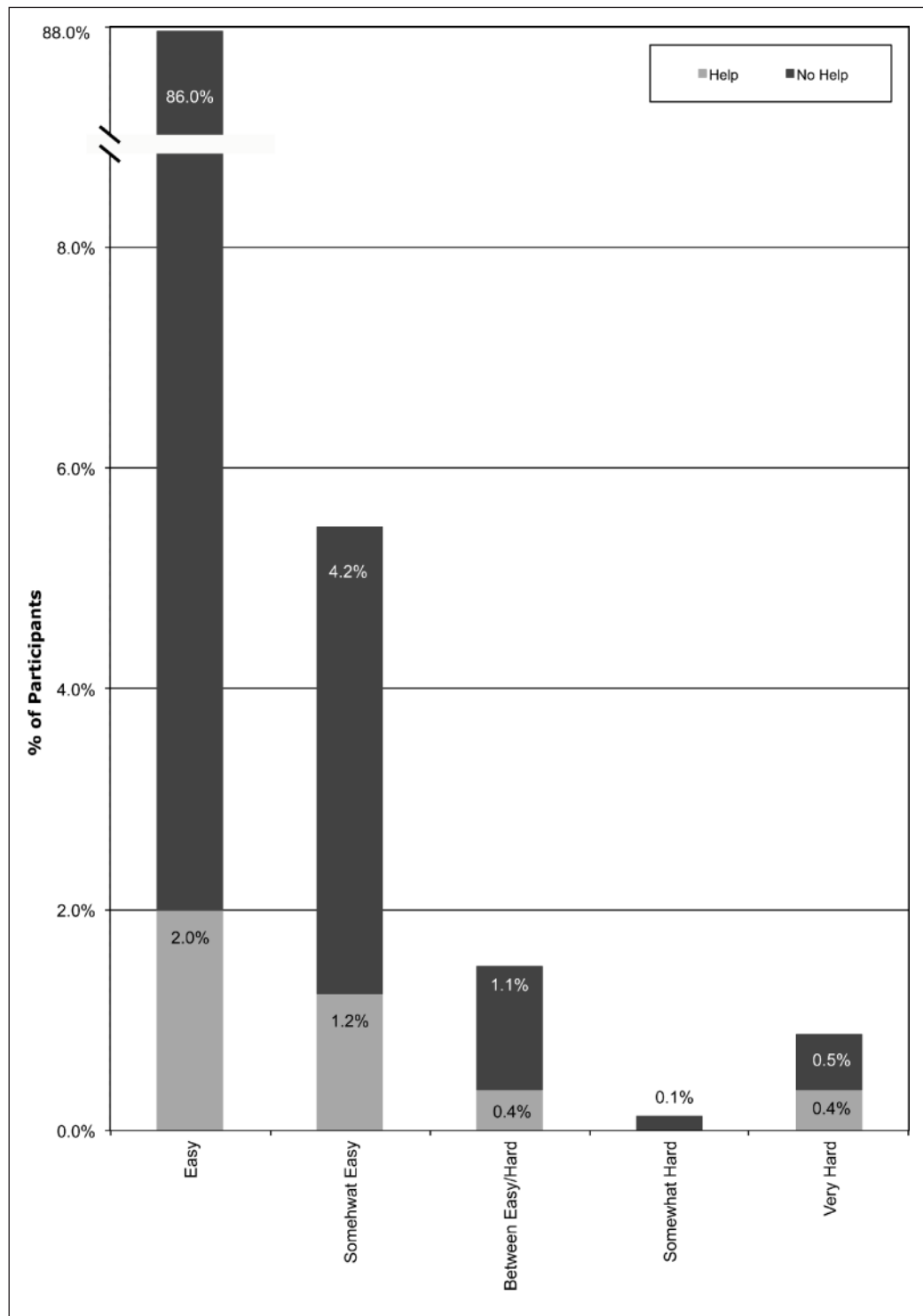


Fig. 1 Self- and observer-reported level of difficulty for the SACAI interview. Self-report answers about perceived difficulty are arrayed as vertical bars from “easy” to “very hard”. Black bar segments represent those who needed minimal or no help from research assistants (n=767); gray bar segments represent those who needed at least moderate help from research assistants (n = 38).* (*Total n = 805 reflects missing data on 36 subjects where interviews were terminated before research assistants were able to report whether the participant required assistance.)

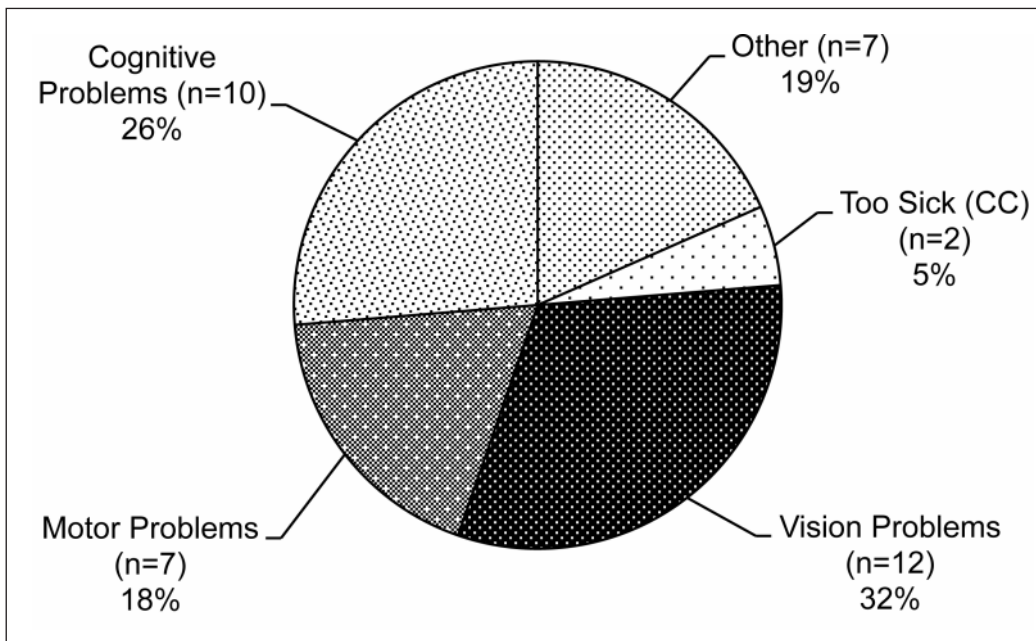


Fig. 2 Research assistant-reported reasons for patients' requiring any assistance (n = 38).

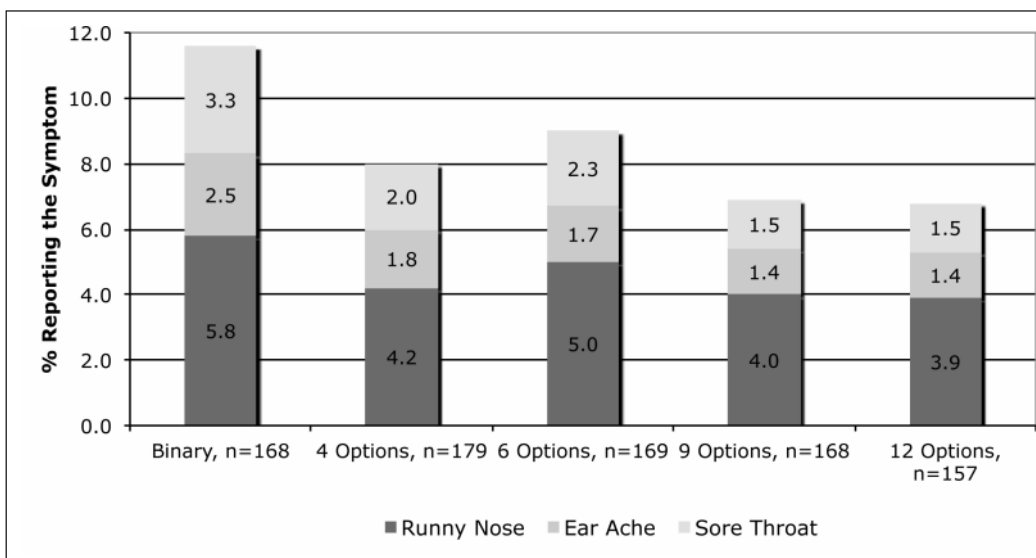


Fig. 3 Prevalence of three targeted symptoms by randomly-assigned question and answer format ('yes/no' question versus 4, 6, 9, or 12 on-screen choices). Patients were asked review of systems questions about the presence of "runny nose," "sore throat," or "earache" presented in one of five possible on-screen formats, assigned at random. Shown are the proportions reporting the presence of each of these symptoms arrayed by the number of response options presented on the screen. Patients were more likely to report each of the three symptoms when presented in binary (yes-no) format.

Table 1 Demographic characteristics of participants

Completed Screening Survey (n = 841)*	
Age (y), mean (range)	41 (18–95)
No. (%) female	493 (59)
No. (%) white	293 (35)
No. (%) black	478 (57)
No. (%) other race	34 (4)
No. (%) refuse to answer race	36 (4)
No. (%) Hispanic	31 (4)
No. (%) refuse to answer Ethnicity	35 (4)
No. (%) with English as a second language	33 (4)
Education level attained, median (range)	12th grade (None to 5th grade – multiple graduate degrees)

*There were no statistically significant differences in demographic characteristics between participants across the four randomly-assigned stakes groups (improve your care n = 201; save your life n = 209; get you seen faster n = 219; save you money n = 212)

Table 2 Demographic, clinical, and equipment-based variation in usability ease of use, error, and efficiency

Demographic	Ease of Use			Error			Efficiency		
	N (%)	Considered SACA Difficult; N (%)	Odds Ratio, P value	N (%)	Age Entry Error; N (%)	Odds Ratio, P value	N (%)	Time-per-click (Seconds); Mean ± SD	P value
Sex									
Male	348 (41.38)	9 (2.59)	†	237 (41.29)	3 (1.27)	†	337 (41.25)	9.35 ± 3.93	
Female	493 (58.62)	11 (2.28)	0.96, 0.93	337 (58.71)	8 (2.37)	1.97, 0.34	480 (58.75)	8.23 ± 3.23	<0.001
Total	841 (100.0)	20 (2.40)		574 (100.0)	11 (1.91)		817 (100.0)	8.73 ± 3.56	
Race									
Black	478 (59.38)	12 (2.51)	†	311 (56.14)	6 (1.93)	†	466 (57.04)	9.11 ± 3.59	†
White	293 (36.40)	5 (1.71)	0.83, 0.73	222 (40.07)	4 (1.80)	1.13, 0.85	286 (35.01)	8.05 ± 3.53	<0.001
Other	34 (4.22)	1 (2.94)	1.48, 0.72	21 (3.79)	0 (0.00)	N/A	70 (7.95)	8.95 ± 3.22	0.04
Total	805 (100.0)	18 (2.23)		554 (100.0)	10 (1.80)		822 (100.0)	8.69 ± 3.59	
Age									
18–40 Years	405 (48.44)	4 (0.99)	†	270 (47.12)	3 (1.11)	†	387 (47.66)	7.79 ± 2.53	
41–65 Years	353 (42.22)	12 (3.40)	2.21, 0.19	251 (43.80)	7 (2.79)	1.30, 0.73	349 (42.98)	9.21 ± 3.79	<0.001
>65 Years	78 (9.33)	4 (5.13)	2.91, 0.18	52 (9.08)	1 (1.92)	0.90, 0.93	76 (9.36)	11.11 ± 3.99	
Total	836 (100.0)	20 (2.39)		573 (100.0)	11 (1.92)		812 (100.0)	8.67 ± 3.57	
Internet access at home									
Yes	460 (54.70)	4 (0.87)	0.21, 0.02	315 (45.12)	2 (0.63)	0.09, 0.03	444 (54.35)	7.98 ± 3.07	
No	381 (45.30)	16 (4.19)	†	259 (54.88)	9 (3.47)	†	373 (45.65)	9.62 ± 3.91	0.002
Total	841 (100.0)	20 (2.38)		574 (100.0)	11 (1.92)		817 (100.0)	8.73 ± 3.56	

Table 2 Continued

Demographic	Ease of Use			Error			Efficiency		
	N (%)	Considered SACA1 Difficult; N (%)	Odds Ratio, P value	N (%)	Age Entry Error; N (%)	Odds Ratio, P value	N (%)	Time-per-click (Seconds); Mean ± SD	P value
High school education									
Yes	614 (73.01)	13 (2.16)	1.21, 0.72	415 (72.30)	7 (1.69)	0.97, 0.96	597 (73.07)	8.38 ± 3.40	
No	227 (26.99)	7 (3.18)	†	159 (27.70)	4 (2.52)	†	220 (26.93)	9.67 ± 3.85	0.01
Total	841 (100.0)	20 (2.38)		574 (100.0)	11 (1.92)		817 (100.0)	8.73 ± 3.56	
Pain level									
0–3	154 (18.31)	1 (0.65)	†	110 (19.16)	1 (0.91)	†	151 (18.48)	9.20 ± 3.59	
4–6	248 (29.49)	8 (3.23)	3.81, 0.22	161 (28.05)	4 (2.48)	3.10, 0.33	240 (29.38)	8.16 ± 3.41	0.77
7–10	439 (52.20)	11 (2.57)	3.38, 0.25	303 (52.79)	6 (1.98)	1.54, 0.70	426 (52.14)	8.88 ± 3.61	
Total	841 (100.0)	20 (2.38)		574 (100.0)	11 (1.92)		817 (100.0)	8.73 ± 3.56	
Mode of Answer Selection									
Digitizer Pen	648 (77.05)	5 (2.60)	0.94, 0.91	457 (79.62)	11 (2.41)	N/A	634 (77.60)	8.86 ± 3.63	
Touch Screen	193 (22.95)	15 (2.37)	†	117 (20.38)	0 (0.00)	N/A	183 (22.40)	8.29 ± 3.32	0.26
Total	841 (100.0)	20 (2.38)		574 (100.0)	11 (1.92)		817 (100.0)	8.73 ± 3.56	
† = Reference group									

References

1. Newman-Toker DE, Pronovost PJ. Diagnostic errors the next frontier for patient safety. *JAMA* 2009; 301(10): 1060-1062.
2. Schiff GD, Hasan O, Kim S, Abrams R, Cosby K, Lambert BL, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med* 2009; 169(20): 1881-1887.
3. Kovacs G, Croskerry P. Clinical decision making: an emergency medicine perspective. *AcadEmergMed* 1999; 6(9): 947-952.
4. Schwartz LR, Overton DT. Emergency department complaints: a one-year analysis. *Ann Emerg Med* 1987; 16(8): 857-861.
5. Trautlein JJ, Lambert RL, Miller J. Malpractice in the emergency department--review of 200 cases. *Ann Emerg Med* 1984; 13(9 Pt 1): 709-711.
6. Miller RA. Medical diagnostic decision support systems--past, present, and future: a threaded bibliography and brief commentary. *J Am Med Inform Assoc* 1994; 1(1): 8-27.
7. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Bmj* 2005; 330(7494): 765.
8. Benaroya M, Elinson R, Zarnke K. Patient-directed intelligent and interactive computer medical history-gathering systems: a utility and feasibility study in the emergency department. *Int J Med Inform* 2007; 76(4): 283-288.
9. Hess R, Santucci A, McTigue K, Fischer G, Kapoor W. Patient difficulty using tablet computers to screen in primary care. *J Gen Intern Med* 2008; 23(4): 476-480.
10. Wolford G, Rosenberg SD, Rosenberg HJ, Swartz MS, Butterfield MI, Swanson JW, et al. A clinical trial comparing interviewer and computer-assisted assessment among clients with severe mental illness. *Psychiatr Serv* 2008; 59(7): 769-775.
11. Acheson LS, Zyzanski SJ, Stange KC, Deptowicz A, Wiesner GL. Validation of a self-administered, computerized tool for collecting and displaying the family history of cancer. *J Clin Oncol* 2006; 24(34): 5395-5402.
12. Langhaug LF, Sherr L, Cowan FM. How to improve the validity of sexual behaviour reporting: systematic review of questionnaire delivery modes in developing countries. *Trop Med Int Health* 2010; 15(3): 362-381.
13. Dugaw JE, Jr., Civello K, Chuinard C, Jones GN. Will patients use a computer to give a medical history? *J Fam Pract* 2000; 49(10): 921-923.
14. Porter SC, Silvia MT, Fleisher GR, Kohane IS, Homer CJ, Mandl KD. Parents as direct contributors to the medical record: validation of their electronic input. *Ann Emerg Med* 2000; 35(4): 346-352.
15. Porter SC, Kohane IS, Goldmann DA. Parents as partners in obtaining the medication history. *J Am Med Inform Assoc* 2005; 12(3): 299-305.
16. Porter SC, Manzi SF, Volpe D, Stack AM. Getting the data right: information accuracy in pediatric emergency medicine. *Qual Saf Health Care* 2006; 15(4): 296-301.
17. Porter SC, Forbes P, Manzi S, Kalish LA. Patients providing the answers: narrowing the gap in data quality for emergency care. *Qual Saf Health Care* 2010; 19(5): e34.
18. Fine AM, Kalish LA, Forbes P, Goldmann D, Mandl KD, Porter SC. Parent-driven technology for decision support in pediatric emergency care. *Jt Comm J Qual Patient Saf.* 2009; 35(6): 307-315.
19. Porter SC, Fleisher GR, Kohane IS, Mandl KD. The value of parental report for diagnosis and management of dehydration in the emergency department. *Ann Emerg Med* 2003; 41(2): 196-205.
20. Bourgeois FT, Porter SC, Valim C, Jackson T, Cook EF, Mandl KD. The value of patient self-report for disease surveillance. *J Am Med Inform Assoc* 2007; 14(6): 765-771.
21. Sawyer MA, Lim RB, Wong SY, Cirangle PT, Birkmire-Peters D. Telementored laparoscopic cholecystectomy: a pilot study. *Stud Health Technol Inform* 2000; 70: 302-308.
22. Abbott PA. The effectiveness and clinical usability of a handheld information appliance. *Nurs Res Pract* 2012; 307258.
23. Schwartz RJ, Weiss KM, Buchanan AV. Error control in medical data. *MD Comput* 1985; 2(2): 19-25.
24. Tang P, McDonald C. Electronic health record systems. In: Shortliffe E, Cimino J, editors. *Biomedical Informatics: Computer Applications in Health Care and Biomedicine (Health Informatics)*. New York 2006. p. 466.
25. McCaig L, Burt C. National Hospital Ambulatory Medical Care Survey: 2003 emergency department summary. Hyattsville, Maryland: National Center for Health Statistics, 2005.
26. Zeiss B, Vega D, Schieferdecker I, Neukirchen H, Grabowski J. Applying the ISO 9126 quality model to test specifications. *Kurzbeiträge*: 2007.

27. National Hospital Ambulatory Medical Care Survey:2010 Emergency Department Summary Tables. 2010; Available from: http://www.cdc.gov/nchs/data/ahcd/nhamcs_emergency/2010_ed_web_tables.pdf. Accessed March 27, 2013.
28. The Nielson Company. Six million more seniors using the web than five years ago. 2009 [cited 2010 October]; Available from: http://blog.nielson.com/nielsenwire/online_mobile/six-million-more-seniors-using-the-web-than-five-years-ago.
29. Wells J, Lewis L. Internet Access in U.S. Public Schools and Classrooms: 1994–2005 (NCES 2007-020). Washington, DC: U.S. Department of Education, 2006.
30. Fisher R. Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research* 1993; 20(2): 303-315.