International Journal of Medical Informatics 106 (2017) 1–8

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Research paper

"Think aloud" and "Near live" usability testing of two complex clinical decision support tools

Safiya Richardson[a,*], Rebecca Mishuris[b], Alexander O'Connell[a], David Feldstein[c], Rachel Hess[d], Paul Smith[c], Lauren McCullagh[a], Thomas McGinn[a], Devin Mann[e]

[a] Hofstra Northwell School of Medicine, United States
[b] Boston University, United States
[c] University of Wisconsin School of Medicine and Public Health, United States
[d] University of Utah School of Medicine, United States
[e] New York University School of Medicine, United States

ARTICLE INFO

Keywords:
Usability
Usability testing
User experience
Clinical decision support
Health informatics
Provider adoption
Workflow

ABSTRACT

Objectives: Low provider adoption continues to be a significant barrier to realizing the potential of clinical decision support. "Think Aloud" and "Near Live" usability testing were conducted on two clinical decision support tools. Each was composed of an alert, a clinical prediction rule which estimated risk of either group A *Streptococcus* pharyngitis or pneumonia and an automatic order set based on risk. The objective of this study was to further understanding of the facilitators of usability and to evaluate the types of additional information gained from proceeding to "Near Live" testing after completing "Think Aloud".

Methods: This was a qualitative observational study conducted at a large academic health care system with 12 primary care providers. During "Think Aloud" testing, participants were provided with written clinical scenarios and asked to verbalize their thought process while interacting with the tool. During "Near Live" testing participants interacted with a mock patient. *Morae* usability software was used to record full screen capture and audio during every session. Participant comments were placed into coding categories and analyzed for generalizable themes. Themes were compared across usability methods.

Results: "Think Aloud" and "Near Live" usability testing generated similar themes under the coding categories visibility, workflow, content, understand-ability and navigation. However, they generated significantly different themes under the coding categories usability, practical usefulness and medical usefulness. During both types of testing participants found the tool easier to use when important text was distinct in its appearance, alerts were passive and appropriately timed, content was up to date, language was clear and simple, and each component of the tool included obvious indicators of next steps. Participant comments reflected higher expectations for usability and usefulness during "Near Live" testing. For example, visit aids, such as automatically generated order sets, were felt to be less useful during "Near-Live" testing because they would not be all inclusive for the visit.

Conclusions: These complementary types of usability testing generated unique and generalizable insights. Feedback during "Think Aloud" testing primarily helped to improve the tools' ease of use. The additional feedback from "Near Live" testing, which mimics a real clinical encounter, was helpful for eliciting key barriers and facilitators to provider workflow and adoption.

1. Background

Clinical decision support (CDS) has demonstrated the ability to shape health care provider behavior towards more evidence based clinical practice by improving diagnosis, treatment, and preventative care services [1–6]. CDS is typically integrated into the electronic health record (EHR) and functions to bring key pieces of evidence or best practice guidelines to the point of care. These tools stand to improve the American healthcare system where on average it takes five years for best practice guidelines to become standard practice [7] and

Abbreviations: CDS, clinical decision support; EHR, electronic health record; GAS, group A *Streptococcus*; SUS, System Usability Scale

* Corresponding author.
E-mail addresses: srichard12@northwell.edu (S. Richardson), rebecca.mishuris@bmc.org (R. Mishuris), aoconnel@northwell.edu (A. O'Connell), df2@medicine.wisc.edu (D. Feldstein), rachel.hess@hsc.utah.edu (R. Hess), paul.smith@fammed.wisc.edu (P. Smith), lmccullagh@northwell.edu (L. McCullagh), tmcginn@northwell.edu (T. McGinn), devin.mann@nyumc.org (D. Mann).

http://dx.doi.org/10.1016/j.ijmedinf.2017.06.003
Received 28 November 2016; Received in revised form 16 June 2017; Accepted 20 June 2017

patients received only 55% of recommended care [8].

Low provider adoption, reported at 10 − 20%, continues to be a significant barrier to realizing the potential of CDS [1]. Efficiency, usefulness, information content, user interface, and workflow have been reported by clinicians to be the keys to effective decision support [1]. These are all components of CDS usability studies and are likely large determinants of clinician adoption rates. Usability testing during the development of CDS allows for its iterative improvement in these areas and has been associated with adoption rates as high as 60% [6].

In "Think Aloud" usability testing, participants verbalize their thoughts as they work through scripted tasks in the EHR. "Think Aloud" testing is resource efficient and provides important feedback on CDS functionality and design [9]. "Near Live" usability testing records providers interacting with a patient actor and the CDS tool. This is more resource intensive but simulates a real clinical environment along with the associated time pressure and natural clinical workflow. These types of usability testing complement each other with the former gathering surface level data and the latter providing insights about underlying workflow issues [10]. "Near Live" usability testing is designed to be conducted after "Think Aloud" usability testing has been conducted and lower level usability issues have been addressed.

The objective of this study was to further understand the determinants of usability by analyzing both the "Think Aloud" and "Near Live" usability testing results of two CDS tools for lessons and themes that could be generalizable to all forms of CDS. The secondary objective was to evaluate the types of additional information gained from proceeding to "Near Live" testing after completing "Think Aloud" usability testing. Usability testing of these two CDS tools was done as a part of the development phase of "Integrated Clinical Prediction Rules: Bringing Evidence to Diverse Primary Care Settings (iCPR2)", a randomized controlled trial evaluating the tools' effect on antibiotic ordering [11]. The CDS tools were composed of an alert, a clinical prediction rule estimating risk of either group A *Streptococcus* (GAS) pharyngitis or pneumonia and an automatic order set based on risk.

2. Methods

This was a qualitative observational study done at the University of Wisconsin, a large academic health care center. "Think Aloud" testing was completed with 4 participants. The tool was revised based on these results before "Near Live" testing was conducted with 8 participants. Different participants were recruited for each type of testing, as is typically the case, to minimize the time commitment required from each of these busy health care providers. Both "Think Aloud" and "Near Live" usability testing were completed with successive participants until saturation was demonstrated. As additional participants complete usability testing they often reiterate the insights of those before them at increasing rates until no new themes emerge. We repeated testing until we stopped hearing new insights from participants. The sample sizes are typical for usability studies and research has demonstrated that they are sufficiently large to elicit the vast majority of usability issues [12–14].

The two CDS tools tested used clinical prediction rules, the Centor Score for GAS pharyngitis and the Heckerling Rule for pneumonia, to calculate the patient's risk for either condition. The tools were both built in Epic Systems EHR and use a standard EPIC alert to inform providers when a patient is appropriate for the tool. The tool is triggered by a reason for visit of sore throat, cough, or upper respiratory tract infection. When triggered, the participant is presented with an alert offering the CDS tool upon opening the chart. If accepted, the participant is taken to a calculator with a list of clinical questions, each of which contributes to a total risk score (Fig. 1). This calculator uses simple yes/no buttons for choosing if criteria are met. Temperature and heart rate are automatically populated based on vitals logged in by the medical assistant. After calculator completion, participants are shown a risk score, identifying patients as low, intermediate or high risk as well as offered an automatic order set based on the calculated risk. The automatic order set included antibiotics based on the calculated risk of bacterial infection. The automatic order sets included documentation for progress notes, laboratory orders, prescription orders, diagnoses, patient's instructions and level of service (Fig. 2).

During both types of usability testing all human-computer interactions, including audio and continuous screen capture, were captured using Morae® (TechSmith, Okemos, MI, USA) software. All verbalized thoughts were transcribed verbatim, coded, and analyzed for generalizable themes. Based on a coding scheme previously developed by the study team, all participant comments were coded under usability, visibility, workflow, content, understand-ability, usefulness or navigation and coded for themes [10]. Emergent codes included the splitting of "usefulness" into "medical" or "practical" usefulness. Participants reported demographic data before every session and completed the System Usability Scale (SUS) afterwards [15]. The SUS is a widely used, validated instrument that measures subjective usability [16,17]. Written informed consent was obtained from all participants. The Institutional Review Boards at both institutions approved the research protocol.

2.1. "Think aloud" usability testing

2.1.1. Participants

Primary care providers were volunteers selected to form a convenience sample, primarily based on clinic location and ease of study conduction. Inclusion criteria required that participants worked in Family Medicine, Internal Medicine or Urgent Care offices, spent at least half of their time providing clinical care and were currently using the EHR system in which the CDS was imbedded. Primary care providers were medical doctors, nurse practitioners and physician assistants.

2.1.2. Procedure

The sessions were conducted in a typical clinic office setting. Each participant was presented with a written clinical case describing a patient with low, intermediate or high risk of either GAS pharyngitis or pneumonia. Following a scripted protocol from the interviewer the usability participant was directed to perform different aspects of clinical documentation including opening the chart, entering patient data, creating a progress note, and placing appropriate orders. While interacting with the tool participants were strongly encouraged to think out loud and to verbalize their thought process. After interacting with the tool the participant was asked a few specific questions about general attitudes towards the tool. The duration of each session was between 25 and 45 min.

2.1.3. Data analysis

Video and transcribed audio recordings were reviewed by two independent coders and placed into coding categories identified in work with earlier versions of these CDS tools. [17] Participant statements were coded under both categories if deemed appropriate by both coders. All discrepancies in the coding were resolved by discussion to achieve a consensus.

2.2. "Near live" usability testing

2.2.1. Participants

Eight primary care providers were selected from volunteers to form a convenience sample, primarily based on clinic location and ease of study conduction. The same inclusion criteria were used as in the "Think Aloud" testing.

2.2.2. Procedure

The session was conducted in a clinic office setting. Each participant was asked to interact with a standardized patient, a patient actor who was trained to portray a case of low, intermediate or high risk GAS