Published in final edited form as:

Int J Med Inform. 2017 October ; 106: 1–8. doi:10.1016/j.ijmedinf.2017.06.003.

“Think Aloud” and “Near Live” Usability Testing of Two Complex Clinical Decision Support Tools

Safiya Richardson¹, Rebecca Mishuris², Alexander O’Connell¹, David Feldstein³, Rachel Hess⁴, Paul Smith³, Lauren McCullagh¹, Thomas McGinn¹, and Devin Mann⁵

¹Hofstra Northwell School of Medicine

²Boston University

³University of Wisconsin School of Medicine and Public Health

⁴University of Utah School of Medicine

⁵New York University School of Medicine

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

Corresponding Author: Safiya Richardson srichard12@northwell.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declarations

Ethics approval and consent to participate

Written informed consent was obtained from all participants. The Institutional Review Boards at both institutions approved the research protocol.

Consent for publication

Not applicable

Availability of data and materials

Datasets analyzed during this study are available from the corresponding author on reasonable request.

CONFLICT OF INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Safiya Richardson, Rachel Mishuris and Alexander O’Connell analyzed and interpreted the qualitative data. David Feldstein, Rachel Hess, Paul Smith and Lauren McCullagh conducted usability testing with participants. Safiya Richardson, Thomas McGinn and Devin Mann were major contributors in writing the manuscript. All authors read and approved the final manuscript.

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.

Keywords

usability; usability testing; user experience; clinical decision support; health informatics; provider adoption; workflow

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 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.

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 (Figure 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 (Figure 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, understandability, 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.

"Think Aloud" Usability Testing

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.

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 25–45 minutes.

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.

"Near Live" Usability Testing

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.

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 pharyngitis or pneumonia. Every participant completed at least one case, two completed two cases each. The participant interacted with the patient actor while navigating through the CDS tool. They were asked to think out loud if they had any challenges or positive opinions of the tool. The study staff observed the session to answer questions, troubleshoot the software and to provide the physical exam information. After completing the testing participants were asked about their reaction to specific components of the tool and for suggestions for improvement. The duration of each session was between 25–45 minutes.

Data Analysis—Two independent coders reviewed all video and transcribed audio. Verbalized thoughts and participant actions were categorized under each code and in the sequence they occurred.

RESULTS

Participants were primarily medical doctors, along with one nurse practitioner and one physician assistant. (Table 1) Our sample was 42% female with an average age of 47.1, 18.7 years of post-graduate practice, and 7.7 years of experience using an EHR. Average SUS was 85.6 during “Think Aloud” testing and 81.3 during “Near Live” testing. SUS scores range from 0 to 100, with 100 being a perfect score.[17] An SUS score of 68 is considered average.[18] Those raw scores would correspond to the 95th percentile rank for “Think Aloud” testing and the 91st for “Near Live” testing.

Common Themes

Previously developed codes were explicitly defined before data analysis. Emergent codes were defined during data analysis. (Table 2) Common themes about visibility, workflow, content, understandability and navigation were observed during both the “Think Aloud” and the “Near Live” testing. (Table 3) Visibility increased when important text was colored differently, larger, underlined, and appropriately located. Participant’s comments consistently reflected that passive alerts that triggered at the time of decision making would allow CDS tool use without disturbing their natural workflow. Comments about content reflected the need for all elements of the tool to incorporate up-to-date and specialty specific medical knowledge. Understandability improved with the use of common medical terminology, providing clear instructions and using simple language. Participant’s ability to navigate through the CDS required each tool component to include obvious indications of next steps.

Additional Themes

Participant comments about usability, practical usefulness and medical usefulness differed during “Near Live” testing from those during “Think Aloud” testing. (Table 4)

“Think Aloud” Usability Testing—Participants commented that the tool allowed them to order diagnostic tests, imaging and treatments at the appropriate time. Ease of use

increased when the tool correctly anticipated their needs by providing what they wanted, when and where they wanted it. In terms of practical usefulness, participants commented that the automatic order set, automatically generated documentation and communication with nurses and patients decreased workload and saved time. In terms of medical usefulness, participants commented positively about the tool's ability to assist with quick recall of the risk factors for GAS pharyngitis or pneumonia. They felt that providing relevant, medically accurate information was helpful for decision making.

“Near Live” Usability Testing—Participants also commented positively about tool simplicity and negatively about unnecessary clicks to navigate through the tool, however there were additional comments about the importance of auto-populate fields and the user's ability to easily and permanently silence the tool. Contrary to comments made during “Think Aloud” testing, participants commented frequently that they would not use the automatically generated order set, patient instructions or documentation because it was incomplete. For example, they felt the order set should include viral respiratory testing, urine legionella antigen testing, etc. They felt that the automatically generated documentation wouldn't be useful if it created a note based on just one of the patient's complaints. Participants also commented frequently that they felt the tool would not be useful because they already knew and used, for example, the Centor Score to assess for risk of GAS pharyngitis. The tool needed to provide unknown information or understanding to users to be perceived as medically useful.

Workflow Map

During “Near Live” testing participants were allowed to interact naturally with patient actors. Every participant demonstrated the same workflow in terms of the order of events. (Figure 3) The alert triggered upon opening the chart based on an appropriate reason for visit. However, all participants navigated away from the alert to first take a history, with or without simultaneously documenting this in a progress note. They then navigated to view the vitals and asked for the physical exam information. Then they looked for, or were directed to, the alert which linked them to the CDS tool. After completing the tool participants counselled the patient while placing the appropriate orders.

DISCUSSION

“Think Aloud” and “Near Live” usability testing of these two CDS tools generated unique insights and lessons generalizable to all forms of CDS. Participant commentary was consistent across participants and across the two tools. This is the first study to evaluate generalizable lessons learned from “Think Aloud” and “Near Live” usability testing of complex clinical decision support tools. Previous studies documented usability findings particular to the tools studied and the relative percentages of positive and negative comments attributable to different components of the tool or components of usability.[10, 11, 19, 20]

This study's focus on qualitative insights facilitated the discovery of the fundamental incompatibility between the EHRs trigger capabilities and natural clinical workflow as well as to clarify the added benefit of “Near Live” usability testing. Comments made during “Think Aloud” testing about visibility, content, understand-ability and navigation primarily

helped to improve the tools ease of use, which is a key driver of technology adoption[21]. Participant comments and observed behavior during “Near Live” testing were more helpful for eliciting key barriers and facilitators to provider workflow and adoption.

During “Near Live” testing participants were more critical of the tool’s usefulness, suggesting that the more realistic environment created a more demanding set of user requirements. During “Think Aloud” testing, the presence of the automatically generated order set, documentation and patient information elicited positive commentary about practical usefulness. However, in “Near Live” testing participants commented that unless these visit aids were complete they would not be useful. For example the automatically generated order set would not be useful unless it included every order that might be placed during a visit for symptoms of an upper respiratory illness.

During “Think Aloud” testing the information the tool presented elicited positive commentary based on accuracy however during “Near Live” testing participants felt strongly that only unknown information would be useful in terms of improving medical decision making. During “Near Live” testing participants remarked that although the tool might be accurate and relevant to the visit there would be no need to use it if they already felt comfortable with diagnosing and treating this condition. Notably, we would expect this to limit the potential impact of any form of CDS directed at changing provider behavior around a common medical problem.

This finding presents a particularly important challenge as most forms of CDS are directed at refining clinical practices around common medical problems. However, these are likely to be ones that providers feel most confident managing without assistance. Overconfidence bias is a significant challenge in medicine, leading to diagnostic errors and adverse events. [22] Just as most people rate themselves as better than average drivers[23], doctors are likely to overestimate their medical judgment concerning diagnoses they consider on a daily basis. Although evidence suggests that about half of antibiotics ordered for acute respiratory conditions are inappropriate[24] during “Near Live” testing, most participants deemed the tool as one that might be useful for other less experienced providers but not useful to them.

Commentary under the coding category usability also supported participants evolving concerns as testing increasingly simulated a real clinical environment. New comments about auto-populate and the ability to easily silence the tool emerged. Scores on the SUS scale during both types of testing indicated slightly decreased perceived usability during “Near Live” testing. This small decline is noteworthy considering the tool had been improved based on prior “Think Aloud” testing. These findings point to increasing expectations by users in terms of tool usability and usefulness as it is tested in increasingly real environments.

“Near Live” testing provided clinical workflow maps as participants interacted with the CDS tool as they would during a real clinical encounter. Participants never used the tool when it triggered upon opening the patient chart. They would not use the tool without first completing the history and physical exam. They went back to look for it when they were ready to think about risk stratifying the patient for GAS pharyngitis or pneumonia. During a

real clinical encounter it might be less likely that a provider would make the effort to look for the tool. These findings suggest that the ideal time for the tool to trigger would be after the history and physical exam but before diagnosis or order entry. However, our prior work using these tools showed better adoption when triggered earlier in the provide workflow.[25]

Most CDS tools trigger before the provider is ready to think about the clinical problem. This discrepancy with natural clinical workflow is a fundamental problem in CDS. It reflects the lack of homogeneity in clinical workflow and the discrepancy between CDS and how providers cognitively process their work.[26] Providers are likely to dismiss tools that trigger before or after the decision making process. Providers gather information from the EHR, for example past medical history and vital signs and then later put orders in using the EHR. Thinking typically happens without and outside of the EHR. Effective CDS may need to be passive, and usable and useful enough that providers make the effort to look for it during the decision making process.

These insights are increasingly relevant as the number of EHR alerts and notifications providers must navigate through increase every day. Primary care doctors receive about 77 total notifications per day[27] and spend nearly two hours on the EHR and desk work for every hour of face to face time with their patients[28]. CDS embedded into the EHR now contributes to the increasing complexity of clinical practice. The presence of too much poorly organized information can be as harmful as having too little; distracting users from important, usable information.[29] There is a growing literature about the unintended consequences of CDS and technology induced errors.[30–35] Combining these types of usability testing allows these tools to be rigorously evaluated and adapted to the needs of providers.

This study has several of limitations. Typical for usability studies, participants were a convenience sample of volunteers, rather than a representative sample. Those who volunteered may have had a particular interest in CDS and therefore comments may have been more positive or negative than if participants were randomly selected. During both “Think Aloud” and “Near Live” testing, an interviewer was present while the participant interacted with and commented about the CDS tool. Participants may have changed their behavior and reported observations as a result of being observed (the Hawthorne effect). This testing was done with just one EHR, Epic Systems, which may limit generalizability. However, this is one of the largest EHRs in the United States and testing was done at a total of 10 different clinics. Lastly, both of these forms of usability testing examine provider thoughts and feelings about the CDS tool outside of the natural clinical environment. Providers may have different thoughts and opinions when asked to use the tool during a real clinical encounter with time pressure and increased patient complexity. All of these limitations are inherent to simulated usability studies and represent standard practice.

CONCLUSION

“Think Aloud” and “Near Live” usability testing provide CDS tool designers with complementary insights that when combined provide a more robust understanding of CDS usability. Participant comments made during “Think Aloud” testing about visibility, content,

understand-ability and navigation primarily helped to improve the tools ease of use. Participant comments and observed behavior during “Near Live” testing, which mimics a real clinical encounter, were more helpful for eliciting key barriers and facilitators to provider workflow and adoption.

Participants displayed higher expectations for usability and usefulness during “Near Live” testing. Additionally, observed behavior during “Near Live” testing provided valuable information about workflow integration. These are significant findings, since most CDS usability testing if done does not complete “Near Live” testing before implementation and this likely contributes to its limited ability to provide significant improvements in healthcare delivery. Combining these types of usability testing provides key insights about provider perspectives, which will serve to improve ease of use of CDS, increase provider adoption and help CDS reach its potential to improve health outcomes.

Acknowledgments

Not applicable

Funding

This project was funded by the National Institutes of Health, National institute of Allergy and Infectious Diseases, under grant #5R01 AI108680-03. The funding body had no role in the design of the study or the collection, analysis, or interpretation of data.

List of Abbreviations

CDS	clinical decision support
EHR	electronic health record
GAS	group A <i>Streptococcus</i>
SUS	System Usability Scale

References

- Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, Samsa G, Hasselblad V, Williams JW, Musty MD. Effect of clinical decision-support systems: a systematic review. *Annals of internal medicine*. 2012; 157(1):29–43. [PubMed: 22751758]
- Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. *Archives of internal medicine*. 2003; 163(12):1409–1416. [PubMed: 12824090]
- Bonnabry P, Despont-Gros C, Grauser D, Casez P, Despond M, Pugin D, Rivara-Mangeat C, Koch M, Vial M, Iten A. A risk analysis method to evaluate the impact of a computerized provider order entry system on patient safety. *Journal of the American Medical Informatics Association*. 2008; 15(4):453–460. [PubMed: 18436900]
- Roshanov PS, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, Nieuwlaat R, Souza NM, Beyene J, Van Spall HG. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. *Bmj*. 2013; 346:f657. [PubMed: 23412440]
- Souza NM, Sebaldt RJ, Mackay JA, Prorok JC, Weise-Kelly L, Navarro T, Wilczynski NL, Haynes RB. Computerized clinical decision support systems for primary preventive care: a decision-maker-researcher partnership systematic review of effects on process of care and patient outcomes. *Implement Sci*. 2011; 6(1):87–99. [PubMed: 21824381]

6. McGinn TG, McCullagh L, Kannry J, Knaus M, Sofianou A, Wisnivesky JP, Mann DM. Efficacy of an evidence-based clinical decision support in primary care practices: a randomized clinical trial. *JAMA internal medicine*. 2013; 173(17):1584–1591. [PubMed: 23896675]
7. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, Kerr EA. The quality of health care delivered to adults in the United States. *New England journal of medicine*. 2003; 348(26):2635–2645. [PubMed: 12826639]
8. Lomas J, Sisk JE, Stocking B. From evidence to practice in the United States, the United Kingdom, and Canada. *The Milbank quarterly*. 1993; 71(3):405–410. [PubMed: 8413068]
9. Rose AF, Schnipper JL, Park ER, Poon EG, Li Q, Middleton B. Using qualitative studies to improve the usability of an EMR. *Journal of biomedical informatics*. 2005; 38(1):51–60. [PubMed: 15694885]
10. Li AC, Kannry JL, Kushniruk A, Chrimes D, McGinn TG, Edonyabo D, Mann DM. Integrating usability testing and think-aloud protocol analysis with “near-live” clinical simulations in evaluating clinical decision support. *International journal of medical informatics*. 2012; 81(11): 761–772. [PubMed: 22456088]
11. Feldstein DA, Hess R, McGinn T, Mishuris RG, McCullagh L, Smith PD, Flynn M, Palmisano J, Doros G, Mann D. Design and implementation of electronic health record integrated clinical prediction rules (iCPR): a randomized trial in diverse primary care settings. *Implementation Science*. 2017; 12(1):37. [PubMed: 28292304]
12. Nielsen J. Estimating the number of subjects needed for a thinking aloud test. *International journal of human-computer studies*. 1994; 41(3):385–397.
13. Virzi RA. Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors: The Journal of the Human Factors and Ergonomics Society*. 1992; 34(4):457–468.
14. Nielsen J. How many test users in a usability study. *Nielsen Norman Group*. 2012; 4(06)
15. Borsci S, Federici S, Lauriola M. On the dimensionality of the System Usability Scale: a test of alternative measurement models. *Cognitive processing*. 2009; 10(3):193–197. [PubMed: 19565283]
16. Lewis, JR., Sauro, J. *International Conference on Human Centered Design*. 2009. Springer; 2009. The factor structure of the system usability scale; p. 94-103.
17. Brooke J. SUS-A quick and dirty usability scale. *Usability evaluation in industry*. 1996; 189(194): 4–7.
18. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Intl Journal of Human—Computer Interaction*. 2008; 24(6):574–594.
19. Mann DM, Lin JJ. Increasing efficacy of primary care-based counseling for diabetes prevention: rationale and design of the ADAPT (Avoiding Diabetes Thru Action Plan Targeting) trial. *Implementation Science*. 2012; 7(1):6. [PubMed: 22269066]
20. Press A, McCullagh L, Khan S, Schachter A, Pardo S, McGinn T. Usability testing of a complex clinical decision support tool in the Emergency Department: lessons learned. *JMIR Human Factors*. 2015; 2(2)
21. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*. 2008; 39(2):273–315.
22. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *The American journal of medicine*. 2008; 121(5):S2–S23.
23. Horswill MS, Waylen AE, Tofield MI. Drivers’ Ratings of Different Components of Their Own Driving Skill: A Greater Illusion of Superiority for Skills That Relate to Accident Involvement1. *Journal of Applied Social Psychology*. 2004; 34(1):177–195.
24. Murphy RA. Prevalence of Inappropriate Antibiotic Prescriptions Among US Ambulatory Care Visits, 2010–2011. *Journal of Emergency Medicine*. 2016; 51(3):337.
25. McGinn TG, McCullagh L, Kannry J, et al. Efficacy of an evidence-based clinical decision support in primary care practices: A randomized clinical trial. *JAMA Internal Medicine*. 2013
26. Holman GT, Beasley JW, Karsh BT, Stone JA, Smith PD, Wetterneck TB. The myth of standardized workflow in primary care. *Journal of the American Medical Informatics Association : JAMIA*. 2016; 23(1):29–37. [PubMed: 26335987]

27. Murphy DR, Meyer AN, Russo E, Sittig DF, Wei L, Singh H. The Burden of Inbox Notifications in Commercial Electronic Health Records. *JAMA internal medicine*. 2016; 176(4):559–560. [PubMed: 26974737]
28. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, Westbrook J, Tutty M, Blike G. Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties. *Annals of internal medicine*. 2016
29. Tierney WM. Improving clinical decisions and outcomes with information: a review. *International journal of medical informatics*. 2001; 62(1):1–9. [PubMed: 11340002]
30. Bullard M, Holroyd BR, Meurer DP. How usability of a web-based clinical decision support system has the potential to contribute to adverse medical events. *human-computer interaction*. 2008; 16:17.
31. Campbell EM, Sittig DF, Ash JS, Guappone KP, Dykstra RH. Types of unintended consequences related to computerized provider order entry. *Journal of the American Medical Informatics Association*. 2006; 13(5):547–556. [PubMed: 16799128]
32. Berger RG, Kichak J. Computerized physician order entry: helpful or harmful? *Journal of the American Medical Informatics Association*. 2004; 11(2):100–103. [PubMed: 14633934]
33. Ash JS, Stavri PZ, Kuperman GJ. A consensus statement on considerations for a successful CPOE implementation. *Journal of the American Medical Informatics Association*. 2003; 10(3):229–234. [PubMed: 12626376]
34. Khajouei R, Jaspers M. The impact of CPOE medication systems' design aspects on usability, workflow and medication orders. *Methods of information in medicine*. 2010; 49(1):3. [PubMed: 19582333]
35. Zulman DM, Shah NH, Verghese A. Evolutionary Pressures on the Electronic Health Record: Caring for Complexity. *Jama*. 2016; 316(9):923–924. [PubMed: 27532804]

HIGHLIGHTS

- “Think Aloud” and “Near Live” usability testing generated unique and generalizable insights.
- Feedback during “Think Aloud” testing primarily helped to improve the tools’ ease of use.
- “Near Live” testing was helpful for eliciting key barriers to provider workflow and adoption.
- During “Near Live” testing participants were more critical of the tools’ usefulness.
- Usability testing helps decision support reach its potential to improve health outcomes.

SUMMARY POINTS

Already Known

- Low provider adoption continues to be a significant barrier to realizing the potential of clinical decision support.
- Efficiency, usefulness, information content, user interface, and workflow have been reported by clinicians to be the keys to effective decision support.

This Study Has Added

- “Think Aloud” and “Near Live” usability testing of these two CDS tools generated unique insights and lessons generalizable to all forms of CDS.
- Insights detailed from “Near Live” testing point to a fundamental incompatibility between the electronic health records trigger capabilities and natural clinical workflow.
- During “Near Live” testing participants were more critical of the tools’ usefulness, suggesting that the more realistic environment created a more demanding set of user requirements.

The screenshot shows a web-based application titled "Strep Pharyngitis Risk Scoring Tool - Pharyngitis". At the top, it displays "Time taken: 2327" and a date of "5/18/2016". Below this is a section labeled "Values By" with a "Create Note" button. A dropdown menu is open, showing "Strep Pharyngitis Risk Scoring Tool - Click Close to continue to SmartSet". The main form contains four clinical questions, each with a "1=Yes" and "0=No" button: "History of Fever?", "Cough?", "Tonsillar Exudates?", and "Tender anterior cervical nodes?". The "Cough?" button is currently selected. Below these questions, the "Strep Pharyngitis Risk Score (out of 4)" is displayed as "3". The "Approx Risk of Strep" is shown as "Intermediate (10-19%)". At the bottom, there are buttons for "Restore", "Close" (with a green checkmark), and "Cancel" (with a red X).

Question	1=Yes	0=No
History of Fever?	<input type="button" value="1=Yes"/>	<input type="button" value="0=No"/>
Cough?	<input checked="" type="button" value="1=Yes"/>	<input type="button" value="0=No"/>
Tonsillar Exudates?	<input type="button" value="1=Yes"/>	<input type="button" value="0=No"/>
Tender anterior cervical nodes?	<input type="button" value="1=Yes"/>	<input type="button" value="0=No"/>

Strep Pharyngitis Risk Score (out of 4): 3

Approx Risk of Strep: Intermediate (10-19%)

Buttons: Restore, Close, Cancel

Figure 1.
Clinical Decision Support Tool Calculator

Associate Providers

Pharmacy No Selected Pharmacy

Sore Throat Decision Support Last Reviewed Date: 12/2/2016 3:15 PM

From BestPractice

Decision Support
Sore Throat Decision Support

Strep Pharyngitis Risk Score (out of 4): 4
Risk scores of 4 are very suggestive of strep pharyngitis or another bacterial cause of pharyngitis. Consider treating for strep pharyngitis without further testing.

Action Steps
1. Click "Accept" to open the SmartSet linked below for treatment options and patient education materials.
2. Click "Cancel" to close the window. You may revisit this alert later in the Best Practice section of the navigator.

Information

SmartSet Information

Risk scores of 4 are very suggestive of strep pharyngitis or another bacterial cause of pharyngitis. Consider treating for strep pharyngitis without further testing.

Documentation

Strep Pharyngitis Risk Score 4+

☐ Strep Pharyngitis Risk Score 4+ [Edit](#)

☐ Full Progress Note [Edit](#)

Prescriptions

Penicillins - First-Line Therapy (Adult)

Penicillin remains the treatment of choice because of its proven efficacy and safety, narrow spectrum and low cost

☐ amoxicillin (AMOXIL) 500 MG cap - 1000 mg 1x/d-10d
1,000 mg, Disp-20 cap, R-0, First occurrence now until 12/25/16

☐ penicillin V potassium (VEETID) 500 MG tab - 2x/d-10d
500 mg, Disp-20 tab, R-0, First occurrence now until 12/25/16

[Antibiotics - Penicillin Intolerant withOUT Anaphylaxis \(Adult\)](#)

[Antibiotics - Patients with Immediate/Severe Reactions to Penicillin or Known Cephalosporin Allergies](#)

Diagnosis

Diagnosis

☒ Streptococcus pharyngitis [J02.0] [Details](#)

☐ Pharyngitis [J02.9] [Details](#)

Patient Instructions

Brief Patient Instructions

☐ Brief Patient Instructions [Edit](#)

Healthwise General Instructions

☐ STREP THROAT (ENGLISH) [Edit](#)

Follow-up

Pharyngitis Follow-up

☐ Patient to follow up in 7 days if symptoms do not improve [Details](#)

Level of Service

Office Visit - Established Patient

☐ 99211, Level 1 (5 min) [Details](#)

☐ 99212, Level 2 (10 min) [Details](#)

☐ 99213, Level 3 (15 min) [Details](#)

☐ 99214, Level 4 (25 min) [Details](#)

☐ 99215, Level 5 (40 min) [Details](#)

[Office Visit - Established Patient, Resident with Faculty](#)

[Office Visit - New Patient](#)

[Office Visit - New Patient, Resident with Faculty](#)

XX

[END OF SMARTSET](#)

Figure 2.
Clinical Decision Support Tool Automatic Order Set

Figure 3.
"Near Live" Provider Workflow

Table 1

Provider Demographics and System Usability Scale Scores

Variable	
<i>Demographics</i>	<i>Providers (n=12)</i>
Age - yr	47.1±13.6
Female Sex — no. (%)	5 (42)
Provider Type — no. (%)	
Medical Doctor (Attending/Faculty)	10 (83%)
Nurse Practitioner	1 (8%)
Physician Assistant	1 (8%)
Years of Post-Graduate Practice	18.7±14.2
Years Using Electronic Health Record	7.7±2.3
<i>System Usability Scale Scores</i>	
“Think Aloud” Usability Testing (n=4)	85.6±24
“Near Live” Usability Testing (n=8)	81.3±14

Table 2

Coding Categories and Definitions

Coding Category	Definition
Usability	Ease of use of the CDS tool, ability to use with minimal effort, i.e. time, clicks, mental energy
Visibility	Ability to quickly recognize key messages and instructions provided by the CDS tool
Workflow	Ability of the CDS tool to fit into the natural order of events in a patient encounter
Content	Medical accuracy or appropriateness of CDS tool text included as orders, patient information, documentation
Understand-ability	Ability to quickly comprehend meaning of text, instructions and the purpose of CDS components
Practical Usefulness	Improves speed or decreases workload during encounter
Medical Usefulness	Improves clinical decision making during encounter
Navigation	Ability to move easily through the CDS tool

Table 3

“Think Aloud” and “Near Live” Usability Testing Results, Common Themes

Coding Category	Example Comments	Theme
Visibility	“What jumps out to me is the blue underlined strep pharyngitis risk screening tool.” (<i>Think Aloud</i>) “But the score is way over on the right so I didn’t actually notice what the score was.” (<i>Near Live</i>)	Visibility improves when important text is colored, larger, underlined and appropriately located.
Workflow	“I much prefer to have stuff in the background that doesn’t force me to have hard stops... There may be a whole series of other things I’m dealing with.” (<i>Think Aloud</i>) “I look at the [tool] after I’ve gathered my information and provided what I think to be the most logical and evidence based direction towards my history, exam and decision making.” (<i>Near Live</i>)	Passive alerts that trigger at the time of decision making allow providers to use CDS tools without disturbing their natural workflow.
Content	“I don’t even know that I would give kids cephalexin, because again it tastes horrible.” (<i>Think Aloud</i>) “It was nice to see that there were available alternatives. And the appropriate dosing and the cued up duration and dose that was appropriate for the condition being treated.” (<i>Near Live</i>)	All elements of the tool should incorporate up-to-date and specialty specific medical knowledge.
Understand-ability	“Do they have rales? I don’t quite know if crackles are rales. Maybe I’m going to say no.” (<i>Think Aloud</i>) “Say absence of cough vs. presence of cough. That double negative may set people off.” (<i>Near Live</i>)	Using common medical terminology, providing clear instructions and using simple language improves understandability.
Navigation	“It says accept or accept and stay. I wouldn’t know what the difference is between that.” (<i>Think Aloud</i>) “I think the click close to get to the [automatic order set]... you’re right it’s not totally intuitive.” (<i>Near Live</i>)	CDS tool components should include obvious indications of next steps.

Table 4

“Think Aloud” and “Near Live” Usability Testing Results, Additional Themes

Coding Category	Example Comments	Theme
Usability	<p><i>Think Aloud</i></p> <p>“Is it going to put accept over here. So you don’t have to navigate over to this corner?”</p> <p>“But it looks like I can just go ahead and order an antibiotic, right off of the screen. I’ve got a diagnosis based on that. So that seems very convenient.”</p> <p><i>Near Live</i></p> <p>“Silence alert, I love that button. Stop yelling at me.”</p> <p>“I liked the fact that it actually obtains and pulls in the clinical information that’s discrete that’s available such as the heart rate and the temperature.”</p>	<p><i>Think Aloud</i></p> <p>Usability increases when CDS tools correctly anticipate their needs by providing what they want, when and where they want it.</p> <p><i>Near Live</i></p> <p>Users perceive ease of use when CDS tools auto-populate with required clinical information and alerts can easily be permanently silenced.</p>
Practical Usefulness	<p><i>Think Aloud</i></p> <p>“I like how it brings these things in. You know, when I’m thinking of typing my note I’m going to be typing in all these things. So, now it’s right there.”</p> <p>“What I liked most was the progress note. It looks like that’s going to save me time. I like saving time.”</p> <p><i>Near Live</i></p> <p>“Part of the reason I don’t use a lot of the [automatic order sets] is that they only have certain bits and pieces of what I want. And then it’s more clicks for me.”</p> <p>“Although, I would put some information [in these patient instructions] here about the antibiotic that I prescribed.”</p>	<p><i>Think Aloud</i></p> <p>Automatically generated documentation and communication with nurses and patients decrease workload and saves time.</p> <p><i>Near Live</i></p> <p>To be perceived as practically useful CDS tools must generate visit aids, like automatic order sets, documentation and patient instructions that are complete. Additionally, CDS tools can help to educate patients.</p>
Medical Usefulness	<p><i>Think Aloud</i></p> <p>“I like the calculator, it’s good. I’m thinking of those things so it’s just nice to have it right there... if you’re kind of trying to piece it together it just takes a second.”</p> <p>“I think it was helpful. I think it helps you make better, informed decisions.”</p> <p><i>Near Live</i></p> <p>“I’m an old guy and I think I know what I’m doing so probably I wouldn’t use it.”</p> <p>“I probably don’t mentally have a calculator, a pre-test prediction rule for pneumonia... So I think it would be useful.”</p>	<p><i>Think Aloud</i></p> <p>CDS that provides information that is perceived as relevant and medically accurate is help for decision making.</p> <p><i>Near Live</i></p> <p>CDS must provide unknown information or understanding to users to be perceived as medically useful.</p>