THIEME
OPEN ACCESS

Usability Testing a Potentially Inappropriate Medication Dashboard: A Core Component of the Dashboard Development Process

Regina Richter Lagha[1]   Zachary Burningham[2,3]   Brian C. Sauer[2,3]   Jianwei Leng[2,3]   Celena Peters[2,3]
Tina Huynh[2,3]   Shardool Patel[2,3]   Ahmad S. Halwani[3,4]   B. Josea Kramer[1,5]

[1] Veterans Affairs Greater Los Angeles Medical Center, Geriatric Research Education and Clinical Center, Los Angeles, California, United States
[2] Division of Epidemiology, Department of Internal Medicine, University of Utah, Salt Lake City, Utah, United States
[3] Salt Lake City Veterans Affairs Medical Center, Health Services Research and Development Center, Salt Lake City, Utah, United States
[4] Division of Epidemiology, Department of Hematology, University of Utah, Salt Lake City, Utah, United States
[5] Division of Geriatric Medicine, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, California, United States

Address for correspondence Regina Richter Lagha, PhD, 11301 Wilshire Boulevard, Los Angeles, CA 90073, United States (e-mail: regina.a.richter@gmail.com).

Abstract

Background   With the increased usage of dashboard reporting systems to monitor and track patient panels by clinical users, developers must ensure that the information displays they produce are accurate and intuitive. When evaluating usability of a clinical dashboard among potential end users, developers oftentimes rely on methods such as questionnaires as opposed to other, more time-intensive strategies that incorporate direct observation.

Objectives   Prior to release of the potentially inappropriate medication (PIM) clinical dashboard, designed to facilitate completion of a quality improvement project by clinician scholars enrolled in the Veterans Affairs (VA) workforce development Geriatric Scholars Program (GSP), we evaluated the usability of the system. This article describes the process of usability testing a dashboard reporting system with clinicians using direct observation and think-aloud moderating techniques.

Methods   We developed a structured interview protocol that combines virtual observation, think-aloud moderating techniques, and retrospective questioning of the overall user experience, including use of the System Usability Scale (SUS). Thematic analysis was used to analyze field notes from the interviews of three GSP alumni.

Results   Our structured approach to usability testing identified specific functional problems with the dashboard reporting system that were missed by results from the SUS. Usability testing lead to overall improvements in the intuitive use of the system, increased data transparency, and clarification of the dashboard's purpose.

Conclusion   Reliance solely on questionnaires and surveys at the end stages of dashboard development can mask potential functional problems that will impede proper usage and lead to misinterpretation of results. A structured approach to usability testing in the developmental phase is an important tool for developers of clinician friendly systems for displaying easily digested information and tracking outcomes for the purpose of quality improvement.

Keywords

► quality improvement
► informatics
► user–computer interface
► inappropriate medication
► decision support systems

Background and Significance

Dashboards have become increasingly popular in the clinical setting as a way to amalgamate large amounts of information from different systems into one platform for the purposes of quality improvement (QI), patient population management, and performance monitoring.[1–5] These tools rely on quality indicators, evidence-based clinical practice guidelines, and risk model algorithms to pull real-time data from various sources and compile the information into an accessible, visually intuitive format.

Dashboard developers often prioritize the accuracy of information; it is equally important to establish the ease of use, or usability, of a dashboard.[6] Focusing on how providers interact with a clinical tool interface can uncover challenges to functionality, which can negatively impact user experience and interpretation of information, thereby affecting overall use.[7]

During the evaluation process, many developers in the health care setting rely on practical tools to assess usability that take into consideration the time limitations of clinical users, including heuristic checklists, questionnaires, surveys, interviews, and focus groups.[8–11] While these can expose end user perceptions of usefulness, implementation feasibility, and satisfaction, a usability study involving real-time observation of users completing specified, typical tasks provides added insight into the quality of user interaction with a given application within the intended setting.[12,13]

The Veterans Affairs' (VA) primary care workforce development project, the Geriatric Scholars Program (GSP), supports the integration of geriatrics into primary care.[14,15] Following intensive training, clinicians initiate an evidence-based QI project at their local institutions, under the guidance of a program mentor, to successfully complete the program. Previously, many Geriatric Scholars relied on time-intensive data collection methods, such as chart review, oftentimes leading to stalled projects and ultimately program incompletion. Program leadership reasoned that dashboards could dramatically reduce the amount of time and effort required to gather baseline and trend data. In 2017, our team began development of a suite of dashboards designed to provide ready access to patient-level data for clinicians enrolled in the program. Employing a Plan Do Study Act framework,[16] the dashboard development process involved identifying target geriatric clinical practice guidelines for dashboard realization, engaging subject matter experts, and consulting with clinical experts.[17] Geriatric Scholars are encouraged but not required to use the dashboards to complete their QI projects.

Geriatric Scholars interested in conducting a local QI project on reducing the prescribing of potentially inappropriate medications (PIMs) can access our PIM dashboard—based on the American Geriatrics Society (AGS) Beers Criteria®[18]—to track their prescribing practices. We customized our dashboard to report data only for a scholar's patient panel, identifying all of the scholar's patients actively on a PIM and reporting the historic proportion of PIMs issued by the scholar. The dashboard consists of a provider summary view (see ►Fig. 1), which allows the user to determine overall prescribing performance, and a patient and medication detail view (see ►Fig. 2), which allows the user to identify specific patients in need of intervention and whether they are candidates for deprescribing.

Objectives

We implemented a formal usability assessment involving real-time observation of specific, common tasks to evaluate potential PIM dashboard user ability to navigate the system, generate reports, and interpret information. This case report describes how our usability testing experience critically informed the development of a dashboard reporting system for the completion of QI projects.

Methods

Participants

Usability testers for the PIM dashboard were selected from a small pool of GSP alumni who volunteered to participate. We targeted individuals who had an active primary care patient
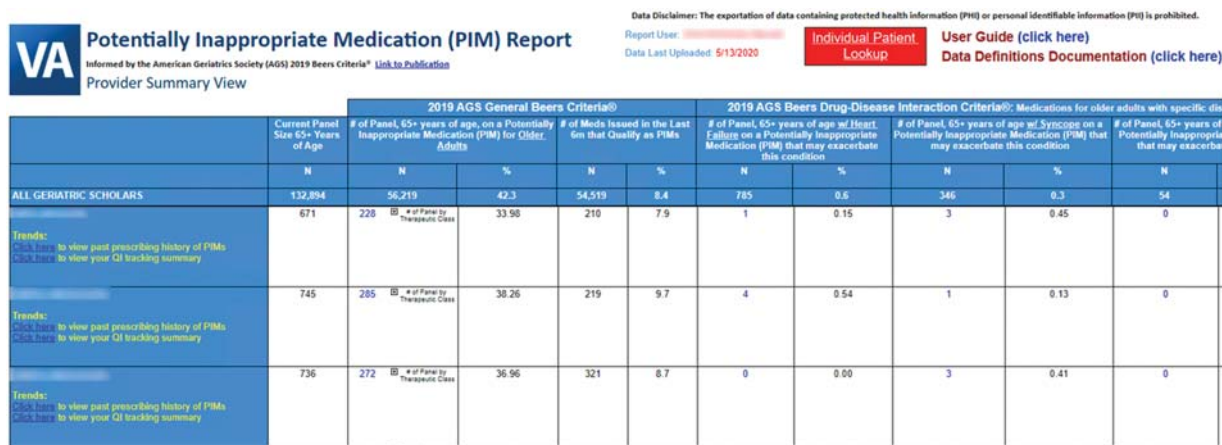

Fig. 1 The potentially inappropriate medication (PIM) dashboard provider summary view supports prescribing performance monitoring at the panel level.
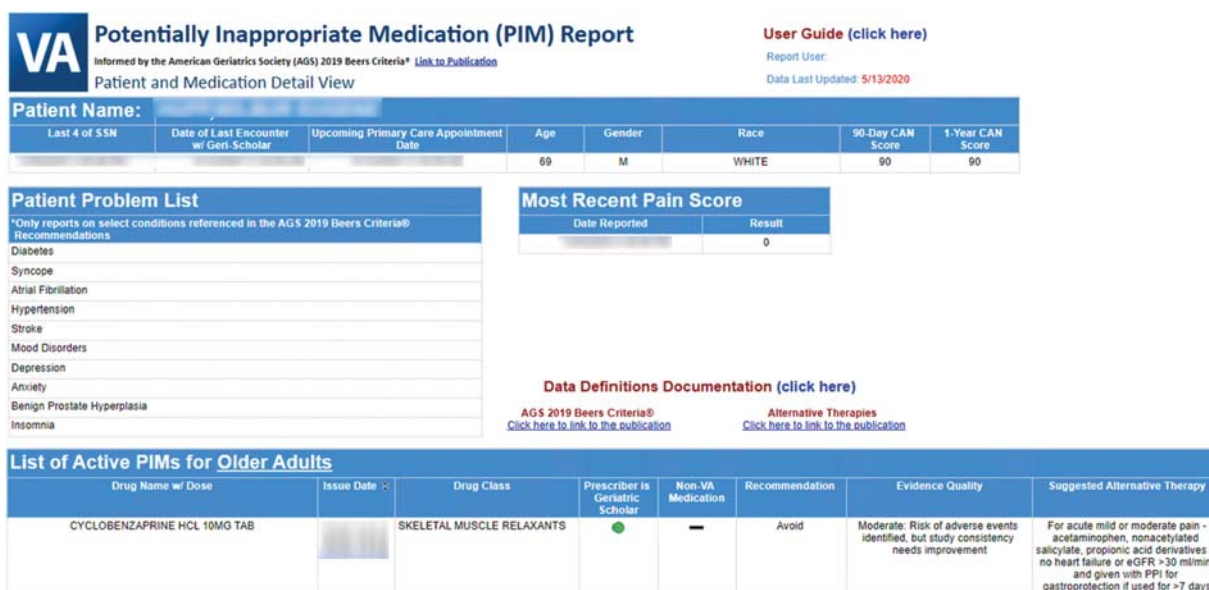
Fig. 2 The potentially inappropriate medication (PIM) dashboard patient and medication detail view enables identification of patients in need of intervention.

Table 1 Inclusion criteria for participants in usability test

Inclusion criteria
GSP alumni
Active primary care patient panel
Patient panel size (65 +) > 100
Proportion of patients (65 +) actively on a PIM > 10%

Abbreviations: GSP, Geriatric Scholars Program; PIM, potentially inappropriate medication.

panel and represented a range of use of contraindicated medications so as to test the dashboard under a variety of different conditions (see ►Table 1). Participants agreed to conduct a 45- to 60-minute recorded telephone interview while sharing their screen. Participants had no prior experience with the dashboard; however, before conducting an interview, we emailed each participant a user manual.

Interview Protocol
Designed to assess both functionality and interpretability, the interview protocol leveraged think-aloud moderating techniques, retrospective questioning about user satisfaction, and administration of the System Usability Scale (SUS).[19] Usability testing relied on both telephone communication and virtual observation through screen-sharing, based on guidance from *The Research Based Web Design and Usability Guidelines* by the U.S. Department of Health and Human Services (HRSA).[20] During the interview, participants were directed, using a structured interview protocol, to complete seven tasks common to the PIM dashboard. Tasks were developed by the team and revised for clarity and simplicity to establish whether dashboard features worked as intended (e.g., back buttons and links) and whether participants could derive appropriate meaning from the

display (e.g., identifying what PIMs a patient is actively on). These tasks were highly specific to the PIM dashboard and allowed the participant to explore the full range of functions available for the purpose of completing a QI project, including collecting baseline data at the panel level, drilling down on specific patients, providing some description of appropriate action to take, and tracking change over time. Participants were then asked more general questions about user satisfaction and completed the SUS.

Commonly employed as a moderating technique when usability testing, think-aloud protocols entail asking participants to speak aloud their mind, including what they are thinking, doing, seeing, or feeling, while they complete a specific task or set of tasks.[21,22] One member of the development team familiar with how to navigate through the dashboard and complete all tasks conducted all aspects of the interview and generated field notes that included both written documentation of participants' observed behaviors and transcription of participants' statements provided by interview audio recordings. A summary of our usability testing approach can be found in ►Fig. 3.

Data Analysis
Two members of the team performed a thematic analysis of the interview field notes to identify emerging themes. Thematic analysis using an inductive approach involves allowing the data to determine themes through a process of coding text, condensing codes into themes, reviewing and revising codes and themes, and finally defining those themes.[23,24] Using this process, these two team members shared and reviewed with the larger development team and their codes to refine and solidify themes. The dashboard development team, which included clinical subject matter experts, informaticians, and programmers, jointly generated strategies to address and improve the dashboard based on these identified challenges.
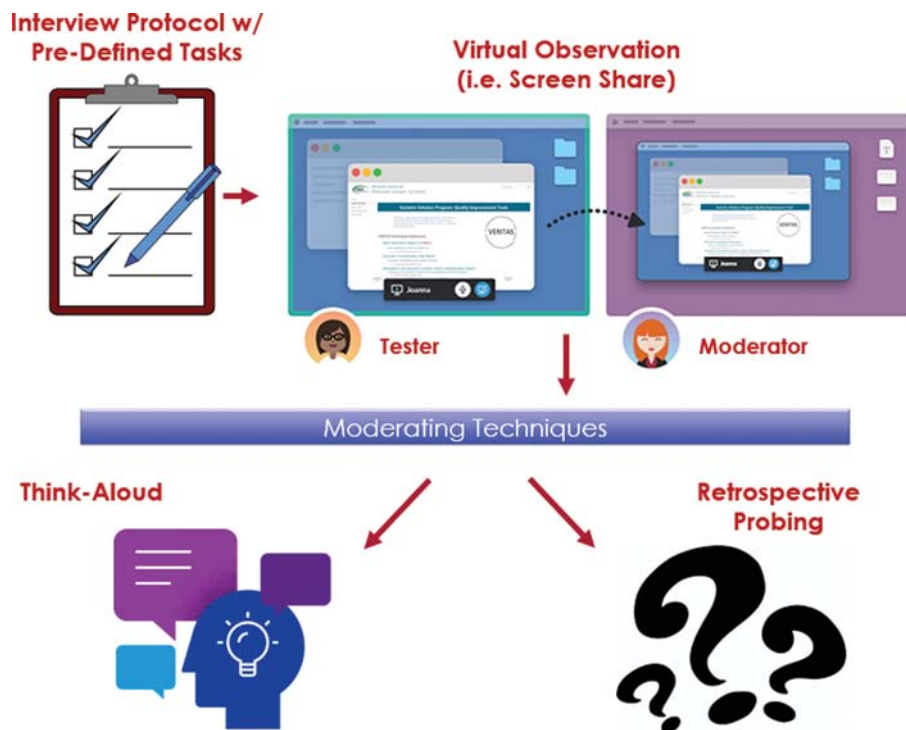
Fig. 3 Graphical summary of our usability testing approach.

Results

Three GSP alumni completed our usability test; each interview took approximately 55 minutes. The mean SUS score of 82.5 (SD = 2.0) indicated above average usability of the PIM dashboard. Analysis of the interview field notes, however, revealed four areas of concern for participants: (1) ease of navigation; (2) interpretability of results; (3) accuracy of results; and (4) perceived intent of the dashboard itself. Each of these concerns resulted in specific changes to the dashboard to facilitate functionality and interpretability of the dashboard.

First, the ability to drill down to more comprehensive data was highly praised—all participants appreciated this feature—but this functionality lacked clearly visible and intuitive action buttons, complicating navigation. For instance, all participants struggled when instructed to find information about their past prescribing, despite a visible link on the user page that read, "Click here to view provider's prescribing pattern of PIMs." These findings prompted the dashboard development team to collectively decide upon and implement changes such as changing the link label to orient the descriptive title to that of the user rather than the developer: "Click here to view past prescribing history of PIMs." Furthermore, "Go Back" action buttons were also added to enhance navigation to and from the dashboard's landing page and drill down views.

Participants also had trouble selecting, when asked, a specific time period parameter (30, 90, or 365 days), which allows a user to view prescribing practices for a defined look-back period. After careful consideration, we elected to remove this feature from the dashboard. However, we have added Statistical Process Control (SPC) charts (see ►Fig. 4) with a predefined look-back period of 365 days so that improvement can be appropriately tracked without the need to manipulate a parameter.

Second, the use of symbols, colors, and language to represent clinical indicators and metrics inadvertently distracted and confused participants. Originally, at the patient-level drill down view, red exclamation points were used to indicate a patient's conditions and/or procedures, relative to the AGS Beers Criteria®. However, participants associated the color with danger and the exclamation with emphasis of a condition, believing the system was alerting them to problems to be solved, dangerous health situations, or even medical errors. These red exclamations were changed to green check marks in subsequent versions of the dashboard to avoid unintended user assumptions. Additionally, where possible, hover over action pop-up dialogues were added to explicitly state the meaning of symbols.

Participants also misinterpreted patient information based on assumptions made about the breadth and scope of presented data. Despite clarification in the user manual, participants did not initially understand that the provided patient-level history included information relevant to AGS Beers Criteria® alone, not a comprehensive summary of a patient's history. A more descriptive header was added to clarify the content of this section.

Third, all participants questioned the accuracy of patient data. One participant became concerned that the system was not accurately gathering and displaying patient data because she could not find a record of a medication she considered a PIM—an opioid—that she knew had been prescribed. In fact, this opioid is not a PIM according to the AGS Beers Criteria®.
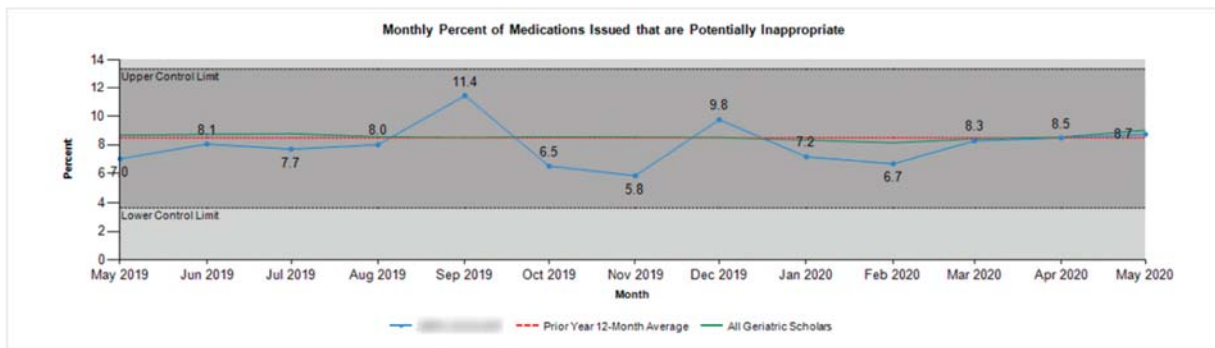
Fig. 4 The Statistical Process Control (SPC) chart monitors prescribing improvement by tracking the percent of potentially inappropriate medications issued by month over 365 days.

The participant questioned not her knowledge of PIMs but rather the dashboard reporting system.

Participants expressed concern on where the data originated from and the repercussions of using those sources. All participants shared the belief that providers can be "lazy" and that reliance on diagnosis codes will lead to inaccurate reports, as providers can fail to code or miscode in the electronic health record (EHR). Because all data sources and transformation methods come with inherent challenges, we decided to be more transparent about from where the dashboard gathered patient information, displaying it clearly within the table, rather than solely in the user manual.

Finally, our last theme concerns the intent of the dashboard. As one participant asked, would the dashboard be used in "monitoring scholars," voicing a concern of many Geriatric Scholars. Within the reports, we do our best to acknowledge the complicated nature of care for older adults. While we reassured all participants that reports were confidential and would not be used to assess prescribing practices by administration, this remains an ongoing concern.

Discussion

Usability testing provides dashboard developers with enhanced insight into how the interface between the clinician user and the report affect the functionality, experience, and overall interpretability of the report. Poor usability can negatively impact not just user experience but also user interpretation of information, potentially threatening patient health.[25,26] Usability methods involving think-aloud protocols can be deemed inappropriate and impractical, when compared with other methods like questionnaires, due to limited clinician time and the oftentimes hectic, interruptive nature of the health care setting.[27] While questionnaires, surveys, interviews, and focus groups can provide information about user satisfaction and perceived ease of use, our experience demonstrates that they can also easily overlook important information about user interpretation of information. This became readily apparent when comparing our above average SUS scores (M = 82.5, SD = 2.0) to our observation field notes; results from the SUS failed to reveal serious functional failures that could lead to misinterpretation and incorrect application of information.

Our experience highlights a potential advantage of combined use of direct observation and think-aloud techniques when considering usability. While the time involved in administering and analyzing data from our usability tests was greater than simply relying on a survey or questionnaire like the SUS, the quality of information was richer.[28] By probing participants for their thoughts during the test, we also gained valuable insight into why a routine task was difficult for the user, enabling us to generate productive solutions to improve the intuitive use of the dashboard.

Analysis of test results highlighted the relationship between functionality and intuitiveness, meaning links and buttons must be clearly labeled in language familiar to the intended clinical user and not to the developer, and properly featured on the page.[29,30] Pop-up dialog boxes can be used to assist in report interpretability. Our experience also demonstrates the importance of transparency of data sources and algorithms.[31] The accuracy of a clinical dashboard reporting system is often contingent on its users. In health care, in principle, all records should be accurate; unfortunately, research shows otherwise.[32] While we cannot control how well a user documents or codes in the EHR, we can be more transparent about the source of our data and our definitions of clinical concepts.

In addition, developers need to consider how confusion over intent might impact the use of the dashboard.[33] The PIM dashboard was developed to assist Geriatric Scholars in completing a mandatory local QI project, not monitor clinical practice. User distrust in the responsible use of this information could negatively impact use of the dashboard. We ourselves continue to refine and focus our messaging.

Limitations

Using think-aloud moderating techniques may have disrupted the natural thought process of our participants.[34,35] Use of this technique also precluded collection of task completion time and other metrics commonly employed in usability testing because it contributed to time on task. Direct observation also increased the amount of time per participant.

Determining the optimal sample for usability testing with clinical end users can be challenging, as developers must

balance the desire to find as many flaws as possible in the functionality of the system with the cost of testing.[7] Though a single end user's experience may not be generalizable, each test contributes to our understanding of the flaws in design that may inhibit proper use. After completing three interviews, we determined that given the capacity of developer and participant time, for our purposes, we had enough information to conclude testing.

Usability testing is just one important component of our dashboard development and evaluation process. While our usability testing involved a small sample of clinicians, assessing other important aspects of the dashboard like utilization and uptake should involve a large base of actual, "real-world" end users. Additionally, evaluation should consider the clinical impact of the dashboard. For instance, was there a greater reduction in the prescribing of PIMs among scholars who used the dashboard compared with those who did not? As we continue to expand our suite of dashboards, formal evaluation may reveal deficiencies in our development process and further inform how we incorporate usability testing in the future.

Conclusion

Usability testing examines whether a dashboard's clinical content is appropriately delivered and easily digested. Direct observation using think-aloud moderating techniques can provide unique, nuanced insight into a clinical user's interaction with a design interface not readily captured by a retrospective satisfaction survey or a quantitative measure like the SUS. Our process shed light not only on functional barriers to the use of the dashboard but also on additional, perhaps less obvious challenges, like perceived accuracy and confusion around the intent of the system. By adopting a structured approach to usability testing, including specifically the use of observation of predetermined tasks vital to system function, developers of clinical tools can improve the design interface, user satisfaction, and functionality, thereby ensuring correct and continual use.

Clinical Relevance Statement

A usability study that focuses on qualities like satisfaction and perceived usefulness can mask specific, easily remedied, functional concerns with a reporting system. Although a usability study that incorporates direct observation may seem impractical, this method can reveal functional concerns with the design interface that impede correct usage of the system, leading to misinterpretation of data. The routine use of these usability testing methodologies by developers of clinical tools has the potential to improve design interface, user satisfaction, as well as effectiveness and functionality for users within the clinical context.

Multiple Choice Questions

1. When conducting an evaluation of clinical tools like dashboard reporting systems, what advantages does a usability study bring over more standard assessment techniques like surveys and questionnaires?

a. Like interviews and focus groups, a usability study allows clinical end users to provide valuable feedback on challenges they face using the system.
b. A usability study can identify potentially hidden, functional challenges for clinical end users interacting with the user-computer interface.
c. There are no real advantages to performing a usability study as standardized assessments and questionnaires are often validated instruments backed by the literature; thus, they are among the best means to evaluate clinical tools designed to enhance quality improvement.
d. A usability study does not provide generalizable results and therefore should be used sparingly when performing an evaluation of clinical tools involving a user–computer interface.

Correct Answer: The correct answer is option b. While standard assessments like surveys and questionnaires as well as focus groups and interviews can expose end user perceptions of usefulness, implementation feasibility, and satisfaction, a structured usability study reveals the quality of user interaction with computer interface, revealing potential functional flaws with the system that can impede its proper usage.

2. When developing a dashboard reporting system, what are the potential consequences of releasing a dashboard reporting system that has not undergone usability testing?

a. There are no consequences. As long as the information being displayed is accurate, that is all that is necessary to ensure a clinical dashboard can be useful in improving care.
b. Without usability testing, a dashboard reporting system cannot offer any utility whatsoever to its intended end user audience.
c. Releasing a dashboard without conducting a usability study can gloss over functional concerns with the system that can impede use.
d. Usability testing ensures the accuracy and functionality of a dashboard reporting system.

Correct Answer: The correct answer is option c. Clinical users may face unexpected functional challenges using the dashboard, leading to misinterpretation of data, mistrust of the system, and eventual disuse.

resources and the use of facilities at the VA SLC Health Care System (HSR&D IDEAS Center), Salt Lake City, UT. The views expressed in this article do not represent the views of the Department of Veterans Affairs.

References

1 Ivers NM, Barrett J. Using report cards and dashboards to drive quality improvement: lessons learnt and lessons still to learn. BMJ Qual Saf 2018;27(06):417–420

2 Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA 1998;280(15): 1339–1346

3 Egan M. Clinical dashboards: impact on workflow, care quality, and patient safety. Crit Care Nurs Q 2006;29(04):354–361

4 Ghazisaeidi M, Safdari R, Torabi M, Mirzaee M, Farzi J, Goodini A. Development of performance dashboards in healthcare sector: key practical issues. Acta Inform Med 2015;23(05):317–321

5 Sujansky W. Heterogeneous database integration in biomedicine. J Biomed Inform 2001;34(04):285–298

6 Armijo D, McDonnell C, Werner K. Electronic Health Record Usability: Evaluation and Use Case Framework. Rockville, MD: Agency for Healthcare Research and Quality; October 2009; AHRQ Publication No. 09(10)-0091-1-EF

7 Bastien JMC. Usability testing: a review of some methodological and technical aspects of the method. Int J Med Inform 2010;79 (04):e18–e23

8 Dolan JE, Lonsdale H, Ahumada LM, et al. Quality initiative using theory of change and visual analytics to improve controlled substance documentation discrepancies in the operating room. Appl Clin Inform 2019;10(03):543–551

9 Dunn S, Sprague AE, Grimshaw JM, et al. A mixed methods evaluation of the maternal-newborn dashboard in Ontario: dashboard attributes, contextual factors, and facilitators and barriers to use: a study protocol. Implement Sci 2016;11:59

10 Lee K, Jung SY, Hwang H, et al. A novel concept for integrating and delivering health information using a comprehensive digital dashboard: an analysis of healthcare professionals' intention to adopt a new system and the trend of its real usage. Int J Med Inform 2017;97:98–108

11 Mlaver E, Schnipper JL, Boxer RB, et al. User-centered collaborative design and development of an inpatient safety dashboard. Jt Comm J Qual Patient Saf 2017;43(12):676–685

12 Bhutkar G, Konkani A, Katre D, Ray GG. A review: healthcare usability evaluation methods. Biomed Instrum Technol 2013: 45–53

13 U.S. Department of Health & Human Services. What & Why of Usability. Available at: https://www.usability.gov/how-to-and-tools/methods/usability-testing.html. Accessed October 12, 2018

14 Kramer BJ, Creekmur B, Howe JL, et al. Veterans Affairs Geriatric Scholars Program: enhancing existing primary care clinician skills in caring for older veterans. J Am Geriatr Soc 2016;64 (11):2343–2348

15 Burningham Z, Chen W, Sauer BC, et al. VA Geriatric Scholars Program's impact on prescribing potentially inappropriate medications. Am J Manag Care 2019;25(09):425–430

16 Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bell D, Reed JE. Systematic review of the application of the plan-do-study-act method to improve quality in healthcare. BMJ Qual Saf 2014;23 (04):290–298

17 2019 American Geriatrics Society Beers Criteria® Update Expert Panel. American Geriatrics Society Health Services Research &

Development (HSR&D) VIReC Clinical Informatics Seminar Series. VA Geriatric Scholars Quality Improvement Dashboards: Leveraging Informatics to Facilitate Change, 6/18/2019. Available at: https://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/video_archive.cfm?SessionID=3643. Accessed July 6, 2020

18 2019 American Geriatrics Society Beers Criteria® Update Expert Panel. American Geriatrics Society 2019 Updated AGS Beers Criteria® for Potentially Inappropriate Medication Use in Older Adults. J Am Geriatr Soc 2019;67:674–694

19 Brooke J. SUS: A "quick and dirty" usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, eds. Usability Evaluation in Industry. London: Taylor and Francis; 1996:189–194

20 Research-Based Web Design & Usability Guidelines. U.S. Department of Health and Human Services. 2006

21 Ericsson K, Simon H. Protocol Analysis: Verbal Reports as Data, 2nd ed. Boston: MIT Press; 1993

22 Romano Bergstrom J. Moderating Usability Tests. U.S. Department of Health and Human Services Usability website. Available at: https://www.usability.gov/get-involved/blog/2013/04/moderating-usability-tests.html. April 2, 2013. Accessed May 16, 2020

23 Castleberry A, Nolen A. Thematic analysis of qualitative research data: Is it as easy as it sounds? Curr Pharm Teach Learn 2018;10 (06):807–815

24 Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017;16(01):1–13

25 Ratwani RM, Reider J, Singh H. A decade of health information technology usability challenges and the path forward. JAMA 2019;321(08):743–744

26 Howe JL, Adams KT, Hettinger AZ, Ratwani RM. Electronic health record usability issues and potential contribution to patient harm. JAMA 2018;319(12):1276–1278

27 Johnson CM, Johnston D, Crowley PK, et al. EHR Usability Toolkit: A Background Report on Usability and Electronic Health Records (Prepared by Westat under Contract No. HHSA 290- 2009–00023I). AHRQ Publication No. 11–0084-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2011

28 Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019;126:95–104

29 Dowding D, Merrill JA. The development of heuristics for evaluation of dashboard visualizations. Appl Clin Inform 2018;9(03): 511–518

30 Dowding D, Merrill JA, Barrón Y, Onorato N, Jonas K, Russell D. Usability evaluation of a dashboard for home care nurses. Comput Inform Nurs 2019;37(01):11–19

31 Voluntary Guidelines for the Design of Clinical Decision Support Software to Assure the Central Role of Healthcare Professionals in Clinical Decision-Making [Internet].Washington, DC: CDS Coalition; 2017. Available at: http://cdscoalition.org/wp-content/uploads/2017/04/CDS-3060-Guidelines-032717-with-memo.pdf. Accessed July 6, 2020

32 Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf 2014;23 (09):727–731

33 Littlejohns P, Wyatt JC, Garvican L. Evaluating computerised health information systems: hard lessons still to be learnt. BMJ 2003;326(7394):860–863

34 Boren MT, Ramey J. Thinking aloud: reconciling theory and practice. IEEE Trans Prof Commun 2000;43(03):261–278

35 Nisbett RE, Wilson TD. Telling more than we know: verbal reports on mental processes. Psychol Rev 1977;84(03):231–259