Review

Work: A Journal of Prevention, Assessment and Rehabilitation
1–13

DOI: 10.1177/10519815251333406
journals.sagepub.com/home/wor
Sage | IOS Press

The most used methods for evaluating health information technology systems usability. A scoping review

Fatemeh Abareshi[1], Farid Zand[2], Roxana Sharifian[3] and Alireza Choobineh[1]

Abstract

Background: Health Information Technology (HIT) systems have become integral to today's health care. However, the success of such systems depends much on their usability because poor usability invites more errors and cognitive loads, ultimately affecting healthcare workers' efficiency and well-being.

Objectives: To identify and summarize the most applied methods of HIT system evaluation, focusing on their impact on work environments and ergonomic practices.

Methods: A scoping review of the literature available from January 2000 up to April 2024 was carried out. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews checklist (PRISMA-ScR) was used. Searches were conducted in four databases. Literature was considered for inclusion if it focused on the HIT, provided information related to the usability of these systems, and provided empirical results of the usability testing. It was written in English.

Results: A total of 591 articles were initially retrieved. Manual searches of reference lists added 21 more articles. Ultimately, 48 articles met the inclusion criteria. The most used method for usability testing was the self-response category.

Conclusions: While qualitative and quantitative methods have been used for usability testing, physiological methods remain underutilized. Eye tracking appeared in only a few studies, and its relationship to usability is still unclear. This study highlights the need to integrate ergonomic principles into HIT system design to enhance usability and support healthcare professional working conditions.

Keywords

usability testing, health information technology, user interface, electronic health record system, user-Centered design

Introduction

Health information technology (HIT) systems are promising means to improve quality, patient safety, and efficiency with the help of some tools in healthcare systems.[1,2] This technology has been widely adopted due to governmental compulsion or incentives such as funding.[3] Despite powerful external forces driving HIT adoption, research has demonstrated that physicians are still dissatisfied with, or resistant to, the technology due to several unintended consequences of problems with design and usability.[3,4] To address usability issues and improve HIT design, usability evaluation research is becoming more prevalent.

Because usability is critical, various methods have been developed and applied to assess the usability of HIT systems in multiple settings. The methods range from objective, performance-based assessment to subjective evaluation through user satisfaction. Commitment toward understanding which usability evaluation methods are

[1]Research Center for Health Sciences, Institute of Health, Shiraz University of Medical Sciences, Shiraz, Iran
[2]Anesthesiology and Critical Care Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
[3]Department of Health Information Management, School of Health Management and Information Sciences, Shiraz University of Medical Sciences, Shiraz, Iran

Corresponding author:
Alireza Choobineh, Research Center for Health Sciences, Institute of Health, Shiraz University of Medical Sciences, Shiraz, Iran.
Email: alrchoobin@sums.ac.ir

Table 1. Keywords.

A HIT	B User Interface	C Usability	D Cognitive workload
Medical Order Entry System	UI	User center design	Efficiency
Computerized Physician Order Entry	User interface	Usability testing	Information overload
Electronic Transmission of Prescription	User-computer interface	usability	Patient safety
Hospital Drug Distribution System	User computer interface	User-center design	Prescribing error
E-Prescribing	Graphical user interface	Ease of use	Medical error
E Prescribing	User interface design	User experience	Workload
CPOE	Human-computer interface	User satisfaction	Workload
Computerized Provider Order Entry System		UX	Human factor
Hospital Medication System			Cognitive workload
Electronic prescribing system			Human factors principle
E-prescription			
Health information system			
CPOE medical system			
Health information technology			
Clinical monitoring technology			
Electronic health record			
Clinical information system			
Electronic medication management system			
Medication alerting system			
Clinical decision support			

most in use and how they are applied within HIT systems is essential to improving the design and deployment of such systems.[3] This scoping review aims to systematically explore the most frequently used for evaluating the usability of HIT, identifying key trends, strengths, limitations, and gaps in the current literature.

Identify the dominating approaches to usability evaluation in HIT systems to inform future research and practices. Guide stakeholders toward more strategic avenues for usability testing that can significantly improve user satisfaction and system efficiency.[5] After all, enhancing HIT systems' usability is essential for optimizing clinical workflow and reducing error risks in healthcare.[6]

Methods

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews checklist (PRISMA-ScR) was used in the study to select the studies and report on the findings.

Information sources and search strategy

An online search was performed for articles published in PubMed, Web of Science, Scopus, and Embase between 2000 and 2024 concerning methods for the usability evaluation of Electronic Health Records. The keywords were divided into four categories: HIT, user interface, usability, and cognitive workload. We used "OR" among the items within each category and "AND" between categories. The selection of the keywords was completed by analyzing previously conducted studies in the field of usability testing methods and through a structured search in the MeSH database for commonly associated terms with the usability of medical records. Keywords are summarized in Table 1.

Study selection

The study selection process is illustrated in the PRISMA flow diagram (Figure 1) and is described as follows:

Identification phase. A comprehensive search was conducted across four major databases: PubMed, Scopus, Web of Science (WOS), and Embase. This search yielded a total of 591 articles. Duplicate records (N = 348) were identified and removed using reference management software, leaving 243 unique articles for further screening.

Screening phase. Articles were screened based on their titles and abstracts. Studies that did not meet the inclusion criteria were excluded. Expressly: Conference Papers (N = 130): Excluded due to insufficient methodological detail and lack of peer review.

Book Sections (N = 14): Excluded as they typically lack the rigor of original research articles. Review Articles (N = 28): Removed to focus on primary studies. Following this screening process, 92 articles were retained for full-text review.

Full-text screening. During this phase, articles were assessed for relevance to the research objectives. Studies were excluded because the study did not evaluate usability methods specific to health information technology systems. The study did not meet language or accessibility requirements.

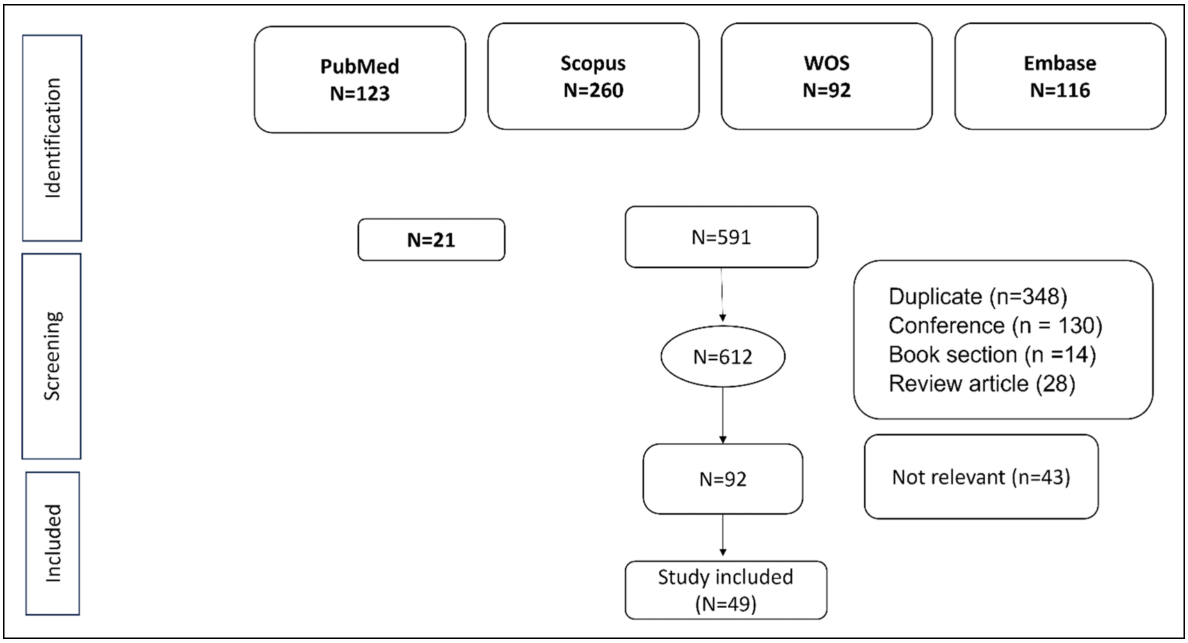

Figure 1. Search process.

After applying these criteria, 43 articles were excluded, leaving 49 for the final review.

Final inclusion. The final set of 49 studies was included in the review, providing a comprehensive overview of usability evaluation methods for health information technology systems. These studies met all inclusion criteria and aligned with the objectives of this scoping review.

Data extraction

The extracted information included the year of publication, country of origin, journal title, type of system, usability testing methods and categories, and name of the standard questionnaire used, along with types of outcome measures.

Synthesis of results

The data was then cleaned and stored within MS Excel for processing, whereby the authors spell-checked and formatted cells accurately. We organized the descriptive data obtained from the included articles through thematic coding into tables and figures; this approach has been very helpful in presenting the review's findings.

Results

Initially, 591 articles were retrieved for this study. Through manual searches of the reference lists, 21 articles have been added. After removing duplicates, conferences, book sections, and review articles, 92 studies were carefully screened for inclusion and exclusion criteria. Finally, the study included 48 articles that met the criteria. The details of the search process and selection of studies are presented visually in Figure1. It is important to note that 4 of the 48 papers from the same project[7,8] were included. Of these, one paper investigates the relationship between electronic health record (EHR) use and the performance, satisfaction, and workload of intensive care unit (ICU) physicians, with a specific focus on identifying systematic differences by gender,[7] and the other utilized eye tracker to capture pupil dilation and gaze fixation to analyze participants interactions with EHR and identifying usability issues.[8] All selected citations evaluated the usability of a type of HIT using different methods. Among the 48 papers, 9 discussed EHR usability evaluations[7–15] while 4 addressed computerized decision support systems (CDSS).[16–19] There were three independent studies examining the usability of the Computerized Physician Order Entry (CPOE),[20–22] Electronic Prescribing System (EPS),[23–25] Nurse Information System (NIS),[26–28] and Hospital Information System (HIS).[29–31] The usability of Picture Archiving and Communication System (PACS),[32,33] Alert interface,[34,35] SIB (Integrated health system),[36,37] and Electronic Audit & Feedback System[38,39] were evaluated separately in two selected citations. Table 2 summarizes the selected articles in detail.

According to Figure 2, most of the publications included in this study were conducted in the USA (n = 20) and Iran (n = 15). As shown in Figure 3, most articles focusing on the most used methods for evaluating the usability of HIT systems were published in 2016 (n = 8), 2018,2019, and 2022 (n = 6), respectively. Most articles (n = 10) have been published in the International

Table 2. Characteristics of the included studies.

Ref	Country	Target population	System Type	Sample size	Method category
9	USA	Physicians	EHR	15	Behavioral
20	USA	Physicians	CPOE*	10	Behavioral
40	Iran	All type users	CPOE	67	Self-report
23	Iran	physicians	EPS Ω	89	Self-report
41	Germany	Anesthesiologists, ICU nurses, respiratory therapists	Remote patient monitoring	10	Self-report, Behavioral
42	USA	Emergency medicine providers	Prescription Drug Monitoring	17	Self-report, Discussion-based, Behavioral
37	Iran	SIB users	SIB**	309	Self-report
43	Iran	-	RIS ∫	-	Expert opinion
10	USA	Physicians	EHR	14	Self-report, Behavioral
26	USA	Nurse	NIS β	12	Self-report
34	USA	Physician, Nurse, Clinical pharmacists	Alert interface	20	Self-report, Discussion-based, Behavioral
44	Sweden	-	virtual health record	-	Expert opinion
45	UK	Patient	MDIQ¥	15	Self-report Discussion-based
11	USA	Physicians	EHR	34	Behavioral
46	Iran	-	LIS€	-	Expert opinion
47	Iran	Physician	EPS	111	Self-report
29	Iran	-	HIS∞	-	Expert opinion
24	UK	users	EPS	20	Self-report, Discussion-based, Behavioral
48	USA	Health professionals	e-GTTЖ	6	Discussion-based
36	Iran	-	SIB	-	Expert opinion
16	USA	-	CDSS θ		Expert opinion
49	USA	Nurse, Practitioners	CDSS	51	Self-report
12	USA	Dental student	Dental EHR	64	Self-report, Discussion-based, Behavioral, Model method
13	USA	Physician, Nurse	EHR	20	Behavioral, Discussion-based
28	Iran	Radiologists, Radiology technician	PAKS ı	230	Self-report
35	USA	Physician, Prescriber	Drug alert interface	12	Behavioral
28	Iran	-	NIS	-	Expert opinion
50	Netherlands	CR professionals	Medi Score CARDSSЉ	7	Self-report, Behavioral
21	Netherlands	-	CPOE	-	Expert opinion
17	USA	Physicians	CDSS	14	Self-report Discussion-based, Behavioral
22	Netherlands	Physicians	CPOE	10	Self-report, Behavioral
18	Netherlands	Physicians	CDSS	13	Self-report, Discussion-based, Behavioral
38	UK	Physicians	e-A&F systems	7	Self-report, Discussion-based, Behavioral

(continued)

Table 2. Continued.

Ref	Country	Target population	System Type	Sample size	Method category
19	USA	Physicians	CDSS		Self-report, Discussion-based
14	China	-	EHR	-	Expert opinion, Model method
30	Iran		HIS		Expert opinion
31	Iran	senior bachelor students	ADT $^{\Psi}$ module of HIS	8	Self-report
33	Iran	Radiologist, Physician	PAKS	8	Self-report, Behavioral
51	UK	Patient	Patient portals for chronic kidney disease	20	Behavioral
7	USA	Physicians	EHR	25	Self-report, Behavioral
8	USA	Physicians	EHR	25	Self-report, Behavioral, Discussion-based
25	Finland	Physicians	EPS	42	Discussion-based
39	UK	-	e-A&F^{ꭓ} systems	-	Expert opinion
52	USA	Clinicians and supporting staff	Electronic medical consultation order forms.	81	Self-report, Discussion-based
15	USA	Physician	EHR	20	Behavioral

*Computerized Physician Order Entry, **national integrated health system, ʃ Radiology Information Systems, β Nurse Information system, ¥ My Diabetes IQ, ₤ Laboratory Information System, Ω Electronic Prescribing System, € Laboratory Information System, ∞ Hospital Information System, θ Computerized Decision Support System, ı Picture Archiving and Communication System, ψ Admission, Discharge, Transfer, ꭓ Electronic Audit& Feedback, Ж Electronic Global Trigger Tool, Љ Cardiac Rehabilitation Computerized Decision Support.

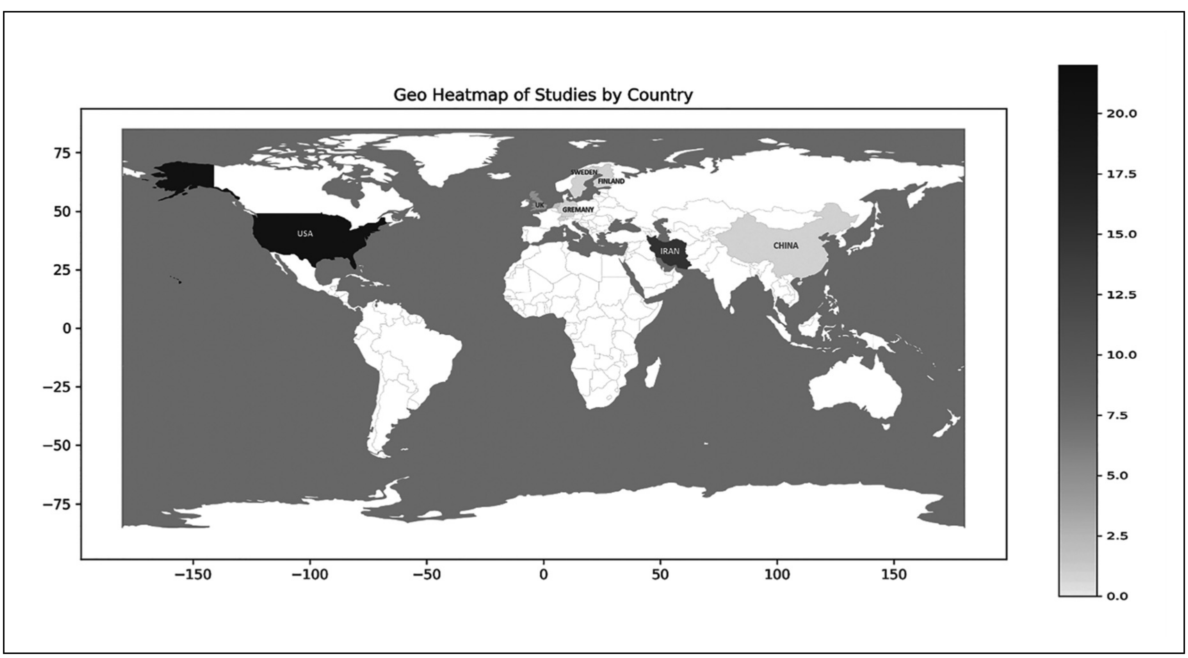

Figure 2. Study distribution by country.

Journal of Medical Informatics. Additionally, physicians were the most common target users in the selected citations. Using only the standard questionnaire to test usability, these systems were reported to have a high sample size, usually more than one hundred,[27,37,47] whereas using other methods resulted in a maximum sample size of 34.

The self-report method, including questionnaires and "think aloud," was the most frequently used method for

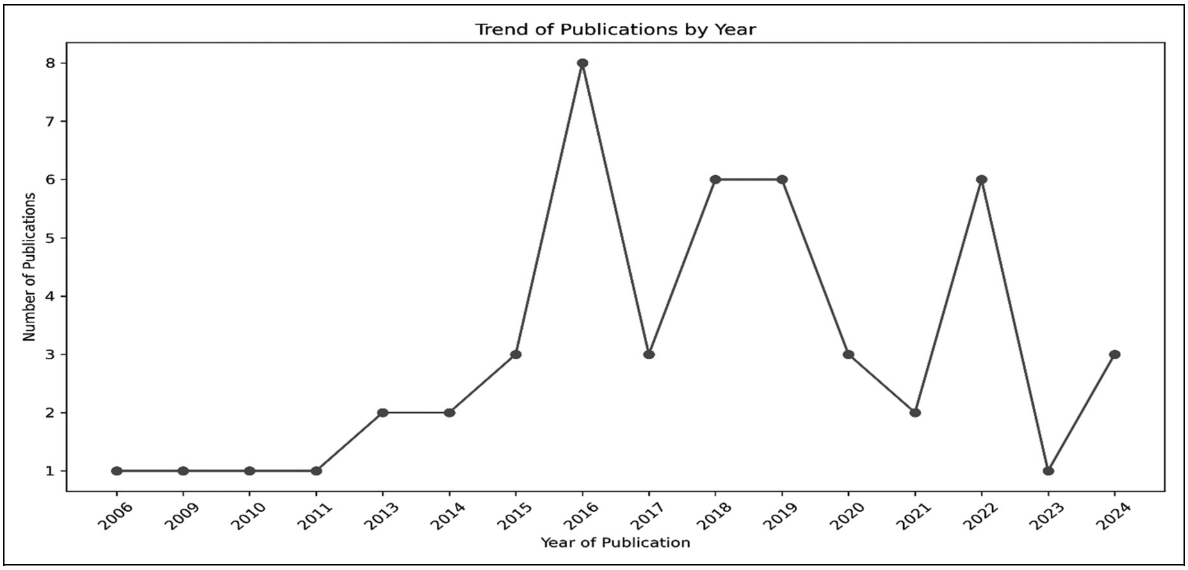

Figure 3. Study distribution by year.

evaluating HIT usability, both individually (14 publications) and in conjunction with other methods (17 publications). The expert opinion category appeared in 9 publications individually and in 4 articles in conjunction with self-report, model method, and discussion-based categories. Behavioral categories were found alone in five publications, but in seven publications, they were well-complemented by self-report and discussion-based methodologies. In addition, discussion methods such as interviews and focus groups are typically employed in combination with other categories (15 publications). Just two articles were found to use the model method category. As a whole, self-reporting accounted for 40% of publications, while behavioral, discussion-based, expert opinion and model methods accounted for 21%, 19%, 17%, and 3%.

In the self-report category, questionnaires were used more than the "think aloud" technique. Five standard questionnaires were identified to evaluate the usability of HIT. SUS (System Usability Scale) was the most used questionnaire (9 publications). CSUQ (Computer System Usability Questionnaire) and QUIS (Questionnaire for User Interface Satisfaction) were used in 4 publications while SUMI (Software Usability Measurement Inventory) and PSSUQ (Post-Study System Usability Questionnaire) were reported in one publication only.

Discussion

Key findings

This paper presents a review of the most prevalent and extensively used methods for assessing HIT system usability. The usability evaluation methods are divided into two main categories. Until 2010, most studies used qualitative methods such as self-report, discussion-based, and expert opinion, whereas, after 2010, researchers applied quantitative methods like behavioral ones alongside qualitative ones.

According to Figure 4, self-report categories such as questionnaires and think-aloud protocols, are commonly utilized if usability methods are in five categories. Among these, standardized and researcher-made questionnaires are employed more extensively than the "think-aloud". A standardized questionnaire is designed for repeated administration, typically featuring a consistent set of questions presented in a fixed order and format, with specific guidelines for generating metrics based on respondents' answers.[53] Standardized questionnaires are created to evaluate participants' satisfaction with the perceived usability of products or systems during usability testing or immediately afterward. In some selected citations, two standard questionnaires were used simultaneously. For instance, Mohammadzadeh et al. utilized SUS and QUIS to assess satisfaction with nine specific factors, including interface factors, terminology and system feedback, learning factors, system capabilities, technical manual, online tutorials, multimedia, teleconferencing, and software installation.[40] Moreover, Khairat et al. used these questionnaires to evaluate the usability of EHR.[8] The System Usability Scale (SUS) is the most commonly used standard questionnaire in the selected citations.[8,10,12,31,38,40,41,49] Due to its fewer items and simpler scale, SUS is considered a faster tool for evaluating perceived usability. Furthermore, unlike SUMI and QUIS, no license is required.

The single most valuable usability engineering technique may be thinking aloud. In a thinking-aloud test, the subject speaks out loud continuously while using the system.[54] Think-aloud protocols elicit information about cognitive

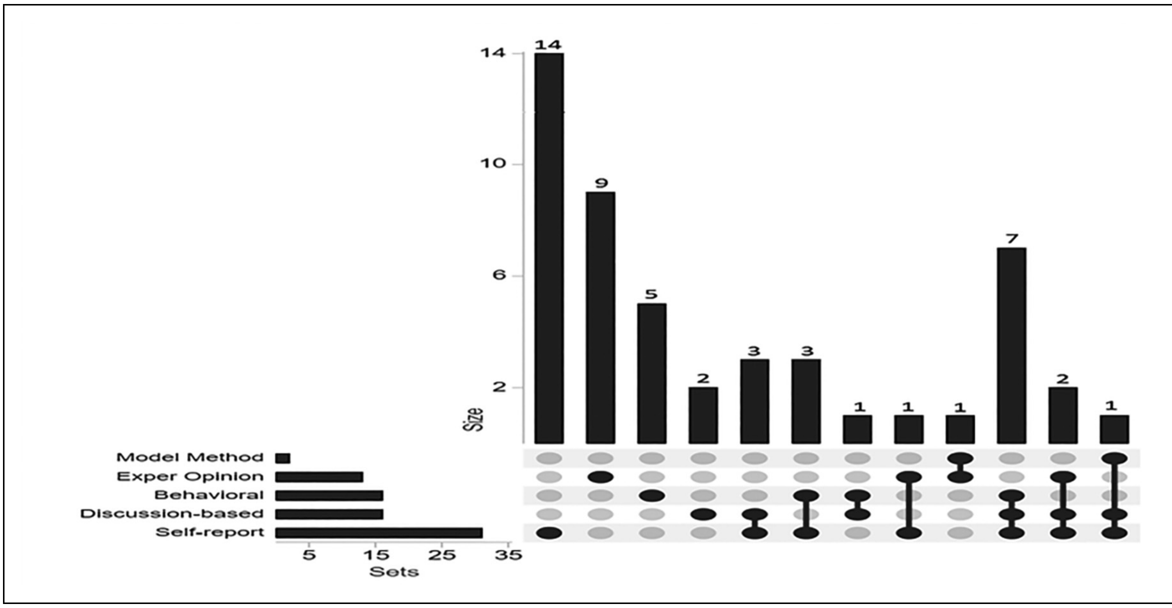

Figure 4. Usability method category.

reasoning during problem-solving.[55] It is a verbal report method formally associated with cognitive psychology.[56] One of its greatest strengths is the qualitative information it can collect from only a few users. Users' comments may also contain vivid and explicit quotes that can enhance the readability and recall of the test report.[57] Besides the questionnaire, this method is the most widely used for evaluating usability. Several studies have used the "think aloud" technique and a standard questionnaire to assess the usability of electronic health records.[10,12,17,18,33,34,41] The selected citations indicated that most researchers employed the "think aloud" technique to collect data under realistic circumstances by creating scenarios or tasks to collect behavioral measures such as task completion time, mouse clicks, keystrokes, errors, number of fixations, or duration.[12,17,18,22,24,33,34,50,58] The selected citations have varying types and numbers of tasks, from one to 19. Poon et al. designed four tasks from more complex to computer-based tasks in the emergency department to evaluate the usability of the prescription drug monitoring program.[42] In a mixed method design, Khairat et al. designed 19 tasks for 4 cases to evaluate the functionality of EHR in the ICU setting.[8]

Interviews, a widely used discussion-based method, are frequently employed alone or in combination with other techniques.[8,12,17–19,24,25,42,48,52,58,59] An interview is ideal for exploratory studies when one does not yet know what one is looking for since the interviewer can adjust it to the situation. There are usually many open-ended questions in interviews, and users are encouraged to elaborate in depth on their answers, often leading to colorful quotes that can be used to illustrate reports or presentations.[57] The disadvantage of interviews is that one cannot necessarily rely on the answers of all participants because when confronted with sensitive questions, people tend to give the answers they think they should, especially when the answers may be embarrassing or socially unacceptable.[57] In contrast, focus groups, another discussion-based method, were rarely mentioned in cited publications.[45] This method requires a relatively large number of representative users. To keep the discussion flowing and have a variety of perspectives represented, focus groups must have a minimum of six participants. Also, conducting more than one focus group session is better since the result of a single meeting may not be representative, and some discussions may divert into minor differences between systems.[57]

The expert opinion category, which includes methods like cognitive walkthroughs and Nielsen's heuristics, is often used independently to identify usability issues in HIT systems.[21,28,29,36,38,43,46,60] Researchers utilized Nielsen's heuristics in some selected publications to categorize the identified usability issues. According to Figure 5, cognitive walkthroughs used less than heuristic techniques in these selected citations. The cognitive walkthroughs are highly structured and explicitly guided by the user's needs. An evaluator, preferably a usability expert, evaluates a user interface by analyzing the cognitive functions required for performing everyday tasks with the assistance of a computer. Because a cognitive walkthrough is intended to help novice users learn a computer application, the evaluator should explore the interface without assistance. When finding severe usability problems, the cognitive walkthrough is more efficient than less severe ones.[61] Moreover, only about one-third of the usability problems detected by heuristic analysis were revealed by cognitive walkthroughs. This is because Cognitive walkthroughs

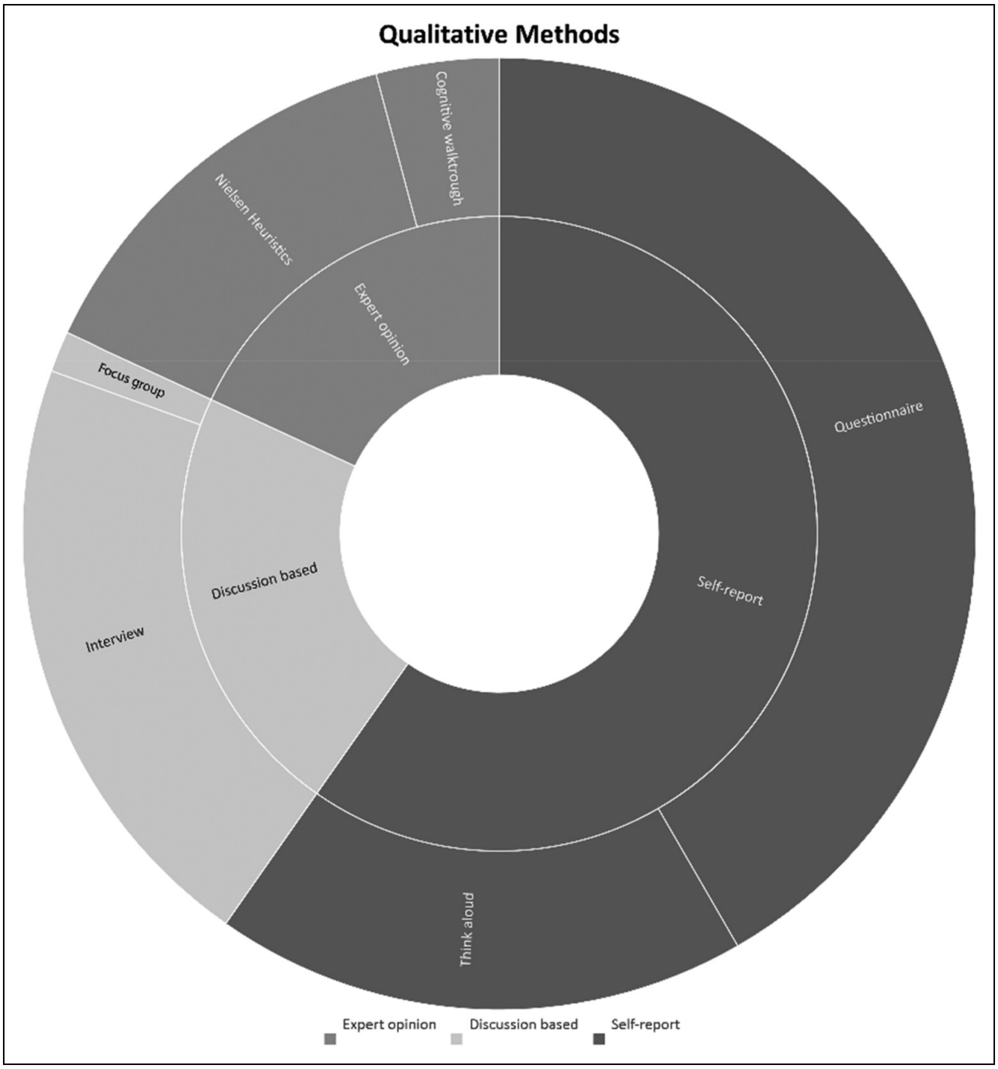

Figure 5. Qualitative methods.

address problems through a structured process, whereas usability principles provide more room to make judgments based on heuristics. In addition to being tedious, the results of cognitive walkthroughs are affected by task description details and a person's background.[61]

In quantitative methods, outcome measures include mouse tracking, mouse click counting, error counting or rate, time on task, and time of successful task completion. To gather these measures, researchers usually use screen capture software like Morae or TURF.[9–12,14,17,18,22,33,34,36,38,42,50,58] Ratwani et al. measured task duration, the number of clicks needed to complete each task, and the accuracy rate among emergency medicine physicians using Electronic Health Records (EHR) for ordering diagnostic imaging, laboratory, and medications.[9] In other studies, the time and number of mouse clicks to complete the given task are considered efficiency.[10,34,42] Efficiency is a usability component mentioned by Neilsen and ISO.[62]

Eye tracking is another behavioral method that researchers have been applying since 2015 to evaluate HIT usability. For example, a study by Schaarup et al. investigated the potential use of the feasibility of eye tracking for assessing the usability of online systems.[63] In another study, Gold examined the assessing EHR using a commercial eye tracker.[11] According to Zahabi, in a review article published in 2015, quantitative methods, such as behavioral and eye tracking, should be used together with qualitative methods when evaluating the usability of electronic health records.[3] The number of fixations and saccades, pupil position, heatmaps, fixation duration, and dwell time are considered in different studies.[8,11,13,24,38,51] Different models of Tobii eye tracker (Tobii Pro X3–120, Tobii X1 Light, Tobii T60, Tobii Pro Glasses) with different characteristics were utilized to collect these measures. Only one study was conducted in a workplace,[33] while the rest were conducted in laboratories or simulated rooms.

Method models are the least commonly used usability evaluation method in the selected citations. In studies, this category is always used in conjunction with other categories. The keystroke level model was applied in two studies.[12,14] A Keystroke-Level Model (KLM) is used in human-computer interaction to predict how long it will take an experienced user to accomplish a routine task using an interactive computer system without error.[64] This technique is a form of GOMS (Goals, Operators, Methods, Selections rules) where a specified set of physical operators is specified, such as keystrokes (typing a diagnosis) and mouse clicks (double-clicking). Further, the model can incorporate mental operators, such as thinking while attempting to locate an item on the interface. There are pre-defined task times for both physical and mental operators (for example, clicking a mouse button takes 0.2 s), which CogTool uses to build an automatic model of expert task performance.[65]

Outcome measures

The measures from the selected citations range from objective methods, such as behavioral category, including task duration, mouse click counting, errors, time on task, and the number of mouse clicks to complete each task in measuring users' performance. These measures are considered efficient and effective in some of the selected citations. For instance, Poncette considered participants' click patterns as efficiency and task completion rate as the effectiveness of a remote patient monitoring system,[41] while in other studies, Poon and Rizvi measured time and mouse clicks to complete the task for the efficiency of systems.[10,42] Farrahi measured effectiveness by the number of completed tasks divided by the total number of tasks,[31] This measure in the study performed by Esfahani was calculated by the number of completed tasks.[33] The NASA-TLX questionnaire assessed the mental workload.[8,10,15,34] The other component of usability was learnability, which was measured using an isometric questionnaire 9241/110 and SUMI questionnaire, according to Haghighi and Jeddi,[23,37] while according to Esfahani, it was measured through the number of tasks that were quickly completed. Farzandipour and Khajouei used CSUQ to measure ease of use.[27,28] This usability component in other studies was measured through an interview with users and through a task difficulty questionnaire with a 7-point Likert scale, ranging from very easy to very difficult.[42,58] The usability errors were assessed according to the number of user errors during task performance,[33] and task completion rate according to the type and number of errors.[38] Usability problems usually derive from "think aloud", expert opinion categories, and interviews.

Gaps and potential for future research

Despite the increasing variety of usability evaluation methods applied in the context of HIT systems and mixed-method approaches, it is noteworthy that electrophysiological approaches such as Electroencephalography (EEG), Electromyography (EMG), Heart rate variability (HRV), and Galvanic Skin Response (GSR) have not been employed in the studies reviewed.

EEG refers to the recording of electrical activity along the scalp that results from firing neurons in the brain. Each state correlates to specific brain wave patterns like gamma, beta, alpha, theta, and delta waves, providing insight into whichever cognitive activity or state is being studied.[66] For example, gamma waves are associated with information processing and memory storage, whereas beta waves are associated with active problem-solving, focus, and concentration. EEG technology has evolved to enable users' states to be analyzed and evaluated in real time, making it especially useful for assessing user engagement and cognitive load to user interaction with health information systems. It is helpful for usability evaluation and designing user interfaces to improve experience and satisfaction.[66]

Galvanic skin response (GSR) measures the skin's electrical conductance, which changes depending on how wet we are from sweat or emotional arousal. This physiological response can be tracked through noninvasive skin sensors, enabling researchers to analyze user reactions to health information systems in real time. GSR measures the conductance of the skin, and changes in GSR can correlate to states of stress or engagement. Thus, it can be crucial for uncovering emotional responses during usability testing. This measure was primarily used to predict stress in gaming and entertainment technology.[67]

Heart rate variability (HRV) is a crucial autonomic nervous system (ANS) metric representing the dynamic interaction of autonomic regulation with the body's innate stress-emotion-behavior "processor." Higher HRV is typically seen as a sign of good health, and lower HRV can indicate stress, anxiety, or problems in the body. The application of HRV in usability testing can reveal how health information systems affect users' physiological stress responses.[68]

Electrophysiological measures can overcome the limitations of traditional approaches for user experience assessment. Such self-reports and/or post-usage questionnaire evaluations cannot yield a more profound understanding of human emotion and the user experience. Many more, researchers say, occur without users' conscious understanding or awareness. They argue that the effects of media consumption on a user's subconscious feelings or psychology can only be determined by utilizing electrophysiological tools and metrics.[69]

In addition, questioning participants about their feelings or experiences interrupts the experience process and exposes their opinions to social biases. The electrophysiological tools will play an important role in removing the social masks to make the final products more functional and bridging the gap of experience aspects (playability, beauty, and aesthetics) that will improve the user experience with interactive systems.[70]

On the other hand, when compared to traditional usability evaluation techniques, electrophysiological approaches have several constraints in terms of cost, complexity, device format, and time investment.[70,71]

Combining physiological signals with traditional usability metrics. For instance, controlled usability testing scenarios could be designed to monitor real-time physiological responses during critical tasks. Analyzing these responses alongside task completion and error rates can provide a more comprehensive picture of system usability. Furthermore, employing machine learning techniques to analyze large-scale physiological data could help identify patterns indicative of usability challenges.

Additionally, while eye-tracking methods have been used in six studies included in this review, there remains a notable gap in analyzing the relationship between quantitative eye-tracking data, such as the number of fixations and fixation duration, and usability outcomes. Since fixation duration indicates task difficulty and cognitive load, future studies could explore its relationship with mental workload as measured by the NASA-TLX questionnaire. Establishing this connection would provide a more comprehensive understanding of how cognitive demands during interaction with HIT are reflected in objective and subjective workload assessment, potentially leading to a more robust usability evaluation. In addition, only one study visualized qualitative eye tracker data like heat maps. No correlation was found between heat maps and usability measures. Reporting technical considerations, such as setting the fixation filter, is important since it is unique for tasks with different content. In one study, this threshold was reported.[51]

Only 18% of usability evaluations resulted in prototype development in the scoping review. In other words, only a few studies have compared the usability of different versions of an HIT interface. Moreover, in most cases, the simulated prototypes were developed with low fidelity, often using computer programs such as Axure, HTML, and JavaScript or Borland JBuilder 2005 software. In one study, an online platform was used for development. For a valid comparison of the usability between the two systems, both prototypes must be developed at the same level of fidelity. Ensuring consistency in fidelity across systems is critical for accurate and meaningful usability comparisons.

In future research, it is recommended that usability evaluation methods for HIT systems consider the distinct perspectives of both developers and end-users. Since these systems are primarily designed by programmers, whose concerns and priorities may differ significantly from those of the actual users, design standards must foster a common language between these two groups. Furthermore, the reviewed studies revealed that ergonomic standards were applied in a limited number of cases. Thus, future research should increase the focus on incorporating ergonomic principles in the design and evaluation of HIT systems to enhance usability.

From an ergonomics perspective, usability evaluation is integral to designing systems that align with human capabilities and limitations. For instance, usability methods that identify cognitive overload can inform the design of interfaces that streamline information presentation, thereby reducing errors and enhancing task performance. This review emphasizes the need for ergonomic usability studies to address the unique challenges faced by healthcare professionals, including multitasking and decision fatigue.

The findings of this review reveal a significant reliance on self-reported usability evaluation methods, which, while valuable for subjective insights, may fall short in capturing nuanced ergonomic factors such as physical strain and task efficiency. Observational methods and task performance metrics, though less frequently utilized, provide a richer dataset for evaluating system usability in alignment with ergonomic principles. These insights suggest that incorporating a mix of methods could enhance the assessment of user-system interaction, leading to more robust system designs.

Conclusion

The main objective of this study was to identify the most used usability methods for evaluating HIT systems. After reviewing related research articles, we found that these methods can be categorized as self-report, discussion-based, behavioral, expert opinion, and model method. According to this review, the most used usability method for evaluating HIT systems was the self-report category, questionnaire, and “think aloud” technique. While these two techniques are fundamental in usability evaluation, each method has its strengths and weaknesses, and there is no single standardized approach. To improve the quality of the evaluation of systems, we recommend that practitioners use a mixed-method approach to usability assessments. This method combines traditional usability metrics (like task success rates, error rates, and completion times) with electrophysiological data such as Electroencephalography (EEG), Heart Rate Variability (HRV), and Galvanic Skin Response (GSR) and behavioral methods like eye tracking. Subjective measures offer researchers and UX practitioners a human perspective, while objective measures provide insights at a fundamental biological level. Also, we stress utilizing real scenarios for evaluation to be better able to allude to the system’s usability in practice. The iterative design cycles, dictated by qualitative and quantitative insights, are critical to improving the human elements of HIT systems through evidence-based use.

Limitations of the study

Our study had some limitations. As noted, although our inclusion criteria were limited to English-language articles, a limitation of this study is that potentially rich sources of research published in other languages may have been excluded, which may have restricted the coverage of perspectives in the studies that we reviewed. We acknowledge that our search was limited to medical databases (PubMed, Web of Science, and Embase) and may not fully capture studies in areas such as human-computer interaction or engineering science. This combination was defined to maintain consistency in the overall goal of the study concerning health information technology (HIT), but we acknowledge that including larger databases will help provide a more comprehensive review in upcoming research. Moreover, we add a note about the exclusion of grey literature (e.g., organizational reports, conference papers, and unpublished studies). Though this method helped ensure a focus on peer-reviewed research, it may have missed insights or practices documented beyond the bounds of academia.

Our data show that science on this topic is mainly focused on the United States and Iran. Although some studies from other countries, e.g., China, Germany, Finland, Sweden, and the United Kingdom, were also found, we restricted the number of countries to those showing the highest number of articles in this area to reflect the current state of research. This disproportionality stems from a ceiling of studies carried out in the mentioned countries present in the selected databases, not from any conscious bias.

Reporting guidelines

This study follows the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist for reporting scoping reviews.

Acknowledgments

We sincerely acknowledge Mr Mehran Ershadipour for assisting with data analysis and designing figures.

ORCID iDs

Fatemeh Abareshi https://orcid.org/0000-0001-9311-4047
Alireza Choobineh https://orcid.org/0000-0002-0866-0404

Statements and declarations

Ethical considerations

Approval was granted by the Ethics Committee of Shiraz University of Medical Sciences (IR.SUMS.REC.1403.524)

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Chaudhry B, Wang J, Wu S, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* 2006; 144: 742–752.
2. Buntin MB, Burke MF, Hoaglin MC, et al. The benefits of health information technology: a review of the recent literature shows predominantly positive results. *Health Aff* 2011; 30: 464–471.
3. Zahabi M, Kaber DB and Swangnetr M. Usability and safety in electronic medical records interface design: a review of recent literature and guideline formulation. *Hum Factors* 2015; 57: 805–834.
4. Clarke MA, Steege LM, Moore JL, et al. (eds). Addressing human computer interaction issues of electronic health record in clinical encounters. In: Design, user experience, and usability health, learning, playing, cultural, and cross-cultural user experience: second international conference, DUXU 2013, held as part of HCI international 2013, Las Vegas, NV, USA, 21–26 July 2013, Proceedings, Part II 2. Springer, 2013.
5. Zheng K, Haftel HM, Hirschl RB, et al. Quantifying the impact of health IT implementations on clinical workflow: a new methodological perspective. *J Am Med Inf Assoc* 2010; 17: 454–461.
6. Ratwani RM, Reider J and Singh H. A decade of health information technology usability challenges and the path forward. *Jama* 2019; 321: 743–744.
7. Khairat S, Coleman C, Ottmar P, et al. Physicians' gender and their use of electronic health records: findings from a mixed-methods usability study. *J Am Med Inf Assoc* 2019; 26: 1505–1514.
8. Khairat S, Coleman C, Newlin T, et al. A mixed-methods evaluation framework for electronic health records usability studies. *J Biomed Inf* 2019; 94: 103175.
9. Ratwani RM, Savage E, Will A, et al. A usability and safety analysis of electronic health records: a multi-center study. *J Am Med Inf Assoc* 2018; 25: 1197–1201.
10. Rizvi RF, Marquard JL, Seyward MA, et al. *Usability evaluation of an EHR's clinical notes interface from the perspective of attending and resident physicians: an exploratory study. MEDINFO 2017: precision healthcare through informatics*. Amsterdam: IOS Press, 2017, pp.1128–1132.
11. Gold JA, Stephenson LE, Gorsuch A, et al. Feasibility of utilizing a commercial eye tracker to assess electronic health record use during patient simulation. *Health Inf J* 2016; 22: 744–757.
12. Walji MF, Kalenderian E, Tran D, et al. Detection and characterization of usability problems in structured data entry interfaces in dentistry. *Int J Med Inf* 2013; 82: 128–138.
13. Wright MC, Dunbar S, Macpherson BC, et al. Toward designing information display to support critical care. *Appl Clin Inf* 2016; 7: 912–929.
14. Xu L, Wen D, Zhang X, et al. Assessing and comparing the usability of Chinese EHRs used in two Peking University hospitals to EHRs used in the US: a method of RUA. *Int J Med Inf* 2016; 89: 32–42.

15. Ahmed A, Chandra S, Herasevich V, et al. The effect of two different electronic health record user interfaces on intensive care provider task load, errors of cognition, and performance. *Crit Care Med* 2011; 39: 1626–1634.
16. Cho H, Keenan G, Madandola OO, et al. Assessing the usability of a clinical decision support system: heuristic evaluation. *JMIR Hum Factors* 2022; 9: e31758.
17. Nguyen KA, Patel H, Haggstrom DA, et al. Utilizing a user-centered approach to develop and assess pharmacogenomic clinical decision support for thiopurine methyltransferase. *BMC Med Inf Decis Making* 2019; 19: 1–13.
18. Kilsdonk E, Peute L, Riezebos RJ, et al. Uncovering health-care practitioners' information processing using the think-aloud method: from paper-based guideline to clinical decision support system. *Int J Med Inf* 2016; 86: 10–19.
19. Casey SD, Reed ME, LeMaster C, et al. Physicians' perceptions of clinical decision support to treat patients with heart failure in the ED. *JAMA Netw Open* 2023; 6: e2344393–e.
20. Pruitt Z, Howe JL, Krevat SA, et al. Development and pilot evaluation of an electronic health record usability and safety self-assessment tool. *JAMIA Open* 2022; 5: ooac070.
21. Khajouei R, de Jongh D and Jaspers MW. Usability evaluation of a computerized physician order entry for medication ordering. In: *Medical informatics in a united and healthy*. Europe: IOS Press, 2009, p.532–536.
22. Khajouei R, Peek N, Wierenga P, et al. Effect of predefined order sets and usability problems on efficiency of computerized medication ordering. *Int J Med Inf* 2010; 79: 690–698.
23. Hayavi-Haghighi MH, Davoodi S, Teshnizi SH, et al. Usability evaluation of electronic prescribing systems from physicians' perspective: a case study from southern Iran. *Inf Med Unlocked* 2024; 45: 101460.
24. Aufegger L, Serou N, Chen S, et al. Evaluating users' experiences of electronic prescribing systems in relation to patient safety: a mixed methods study. *BMC Med Inf Decis Making* 2020; 20: –8.
25. Kauppinen H, Ahonen R, Mäntyselkä P, et al. Medication safety and the usability of electronic prescribing as perceived by physicians—A semistructured interview among primary health care physicians in Finland. *J Eval Clin Pract* 2017; 23: 1187–1194.
26. Rogers ML, Sockolow PS, Bowles KH, et al. Use of a human factors approach to uncover informatics needs of nurses in documentation of care. *Int J Med Inf* 2013; 82: 1068–1074.
27. Khajouei R and Abbasi R. Evaluating nurses' satisfaction with two nursing information systems. *CIN Comput Inf Nurs* 2017; 35: 307–314.
28. Farzandipour M, Nabovati E, Tadayon H, et al. Usability evaluation of a nursing information system by applying cognitive walkthrough method. *Int J Med Inf* 2021; 152: 104459.
29. Atashi A, Khajouei R, Azizi A, et al. User interface problems of a nationwide inpatient information system: a heuristic evaluation. *Appl Clin Inf* 2016; 7: 89–100.
30. Farzandipour M, Nabovati E, Zaeimi G-H, et al. Usability evaluation of three admission and medical records subsystems integrated into nationwide hospital information systems: heuristic evaluation. *Acta Inf Med* 2018; 26: 33.
31. Farrahi R, Rangraz Jeddi F, Nabovati E, et al. The relationship between user interface problems of an admission, discharge and transfer module and usability features: a usability testing method. *BMC Med Inf Dec Making* 2019; 19: 1–8.
32. Farzandipour M, Jabali MS, Nickfarjam AM, et al. Usability evaluation of selected picture archiving and communication systems at the national level: analysis of users' viewpoints. *Int J Med Inf* 2021; 147: 104372.
33. Esfahani MZ, Khajouei R and Baneshi MR. Augmentation of the think-aloud method with users' perspectives for the selection of a picture archiving and communication system. *J Biomed Inf* 2018; 80: 43–51.
34. Russ AL, Zillich AJ, Melton BL, et al. Applying human factors principles to alert design increases efficiency and reduces prescribing errors in a scenario-based simulation. *J Am Med Inf Assoc* 2014; 21: e287–ee96.
35. Xie M, Weinger M, Gregg W, et al. Presenting multiple drug alerts in an ambulatory electronic prescribing system. *Appl Clin Inf* 2014; 5: 334–348.
36. Jeddi FR, Nabovati E, Bigham R, et al. Usability evaluation of a comprehensive national health information system: a heuristic evaluation. *Inf Med Unlocked* 2020; 19: 100332.
37. Jeddi FR, Nabovati E, Bigham R, et al. Usability evaluation of a comprehensive national health information system: relationship of quality components to users' characteristics. *Int J Med Inf* 2020; 133: 104026.
38. Brown B, Balatsoukas P, Williams R, et al. Multi-method laboratory user evaluation of an actionable clinical performance information system: implications for usability and patient safety. *J Biomed Inf* 2018; 77: 62–80.
39. Brown B, Balatsoukas P, Williams R, et al. Interface design recommendations for computerised clinical audit and feedback: hybrid usability evidence from a research-led system. *Int J Med Inf* 2016; 94: 191–206.
40. Mohammadzadeh Z, Nickfarjam AM, Atoof F, et al. Evaluating usability of computerized physician order entry systems: insights from a developing nation. *Inf Med Unlocked* 2024; 47: 101487.
41. Poncette A-S, Mosch LK, Stablo L, et al. A remote patient-monitoring system for intensive care medicine: mixed methods Human-Centered design and usability evaluation. *JMIR Hum Factors* 2022; 9: e30655.
42. Poon SJ, Greenwood-Ericksen MB, Gish RE, et al. Usability of the Massachusetts prescription drug monitoring program in the emergency department: a mixed-methods study. *Acad Emerg Med* 2016; 23: 406–414.
43. Rezaei-Hachesu P, Pesianian E and Mohammadian M. Evaluating usability of radiology information systems in hospitals of Tabriz University of medical sciences. *Acta Inf Med* 2016; 24: 42.
44. Scandurra I, Hägglund M, Moström D, et al. Heuristic evaluation extended by user analysis: a fast and efficient method to

identify potential usability problems in health information systems. *J Inf Technol Healthcare* 2006; 4: 317–325.
45. Shields C, Cunningham SG, Wake DJ, et al. User-centered design of a novel risk prediction behavior change tool augmented with an artificial intelligence engine (MyDiabetesIQ): a sociotechnical systems approach. *JMIR Hum Factors* 2022; 9: e29973.
46. Agharezaei Z, Khajouei R, Ahmadia L, et al. Compliance with design principles: a case study of a widely used laboratory information system. *East Mediterr Health J* 2020; 26: 1456–1464.
47. Alinejhad A, Nourani A and Jebraeily M. Usability evaluation of the user interface in electronic prescribing systems of Iran health insurance organization and social security organization. *J Health Adm* 2022; 25: 78–94.
48. Blakeney EA-R, Dardas T, Zierler BK, et al. Development and usability testing of a system to detect adverse events and medical mistakes. *CIN Comput Inf Nurs* 2023; 41: 330–337.
49. Dodson C and Layman L. Refinement of a pharmacogenomics app for dosing guidelines for oncology: findings from the usability evaluation. *Ann Transl Med* 2022; 10: 1261–1261.
50. van Engen-Verheul MM, Peute LW, de Keizer NF, et al. Optimizing the user interface of a data entry module for an electronic patient record for cardiac rehabilitation: a mixed method usability approach. *Int J oMed Inf* 2016; 87: 15–26.
51. Fraccaro P, Vigo M, Balatsoukas P, et al. Presentation of laboratory test results in patient portals: influence of interface design on risk interpretation and visual search behaviour. *BMC Med Inf Decis Making* 2018; 18: 1–12.
52. Savoy A, Militello LG, Patel H, et al. A cognitive systems engineering design approach to improve the usability of electronic order forms for medical consultation. *J Biomed Inf* 2018; 85: 138–148.
53. Sauro J and Lewis JR. *Quantifying the user experience: practical statistics for user research*. United States: Morgan Kaufmann, 2016.
54. Lewis C. *Using the" thinking-aloud" method in cognitive interface design*. Yorktown Heights, NY: IBM TJ Watson Research Center, 1982.
55. Nielsen J. *Response times: the three important limits*. United States: Usability Engineering, 1993.
56. Simon HA and Ericsson KA. Protocol analysis: Verbal reports as data. (No Title). 1984.
57. Nielsen J. *Usability engineering*. United States: Morgan Kaufmann Publishers In. Inc–1993, 1993.
58. Salz T, Schnall RB, McCabe MS, et al. Incorporating multiple perspectives into the development of an electronic survivorship platform for head and neck cancer. *JCO Clin Cancer Inf* 2018; 2: 1–15.
59. Bowman S. Impact of electronic health record systems on information integrity: quality and safety implications. *Perspect Health Inf Manage* 2013; 10: 1–19.
60. Conforti D, Costanzo D, Lagani V, et al. HEARTFAID: a knowledge based platform for supporting the clinical management of elderly patients with heart failure. *J Inf Technol Healthcare* 2006; 11: 46–51.
61. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. *Int J Med Inf* 2009; 78: 340–353.
62. Usability 101: Introduction to usability [Internet], https://www.nngroup.com/articles/usability-101-introduction-to-usability/ (2012).
63. Schaarup C, Hartvigsen G, Larsen LB, et al. *Assessing the potential use of eye-tracking triangulation for evaluating the usability of an online diabetes exercise system. MEDINFO 2015: eHealth-enabled health*. Amsterdam: IOS Press, 2015, pp.84–88.
64. Card SK, Moran TP and Newell A. The keystroke-level model for user performance time with interactive systems. *Commun ACM* 1980; 23: 396–410.
65. John BE. Information processing and skilled behavior. HCI models, theories, and frameworks. 2003.
66. Alfimtsev A, Basarab M, Devyatkov V, et al. A new methodology of usability testing on the base of the analysis of user's electroencephalogram. *J Comput Sci Appl* 2015; 3: 105–111.
67. Laufer L and Németh B. Your skin knows when you will jump. In: *Proceedings of the 14th European conference on cognitive ergonomics: invent! explore!*. London, United Kingdom: Association for Computing Machinery, 2007, p.277–280.
68. Zaki T and Islam MN. Neurological and physiological measures to evaluate the usability and user-experience (UX) of information systems: a systematic literature review. *Comput Sci Rev* 2021; 40: 100375.
69. Taffese T. A review of using EEG and EMG psychophysiological measurements in user experience research. 2017.
70. Ganglbauer E, Schrammel J, Deutsch S, et al. (eds). Applying psychophysiological methods for measuring user experience: possibilities, challenges and feasibility. In: *Workshop on user experience evaluation methods in product development*. New Jersey: Citeseer, 2009, pp.1–4.
71. Kivikangas JM, Chanel G, Cowley B, et al. A review of the use of psychophysiological methods in game research. *J Gaming Virtual Worlds* 2011; 3: 181–199.