Journal of the American Medical Informatics Association, 2024, **31(1)**, 240–255
https://doi.org/10.1093/jamia/ocad188
Advance access publication 22 September 2023
Review

Review

The relationship between electronic health records user interface features and data quality of patient clinical information: an integrative review

Olatunde O. Madandola, MPH, RN[1,*], Ragnhildur I. Bjarnadottir, PhD, MPH, RN[1], Yingwei Yao, PhD[1], Margaret Ansell, MLIS, AHIP[2], Fabiana Dos Santos, PhD, MSN, RN[1], Hwayoung Cho, PhD, RN[1], Karen Dunn Lopez, PhD, MPH, RN, FAAN[3], Tamara G.R. Macieira, PhD, RN[1], Gail M. Keenan, PhD, RN, FAAN[1]

[1]University of Florida College of Nursing, Gainesville, FL, United States, [2]University of Florida Health Sciences Library, Gainesville, FL, United States, [3]University of Iowa College of Nursing, Iowa City, IA, United States

*Corresponding author: Olatunde O. Madandola, MPH, RN, University of Florida College of Nursing, 1225 Center Dr., Gainesville, FL 32610-0197 (omadandola@ufl.edu)

Abstract

Objectives: Electronic health records (EHRs) user interfaces (UI) designed for data entry can potentially impact the quality of patient information captured in the EHRs. This review identified and synthesized the literature evidence about the relationship of UI features in EHRs on data quality (DQ).

Materials and methods: We performed an integrative review of research studies by conducting a structured search in 5 databases completed on October 10, 2022. We applied Whittemore & Knafl's methodology to identify literature, extract, and synthesize information, iteratively. We adapted Kmet et al appraisal tool for the quality assessment of the evidence. The research protocol was registered with PROSPERO (CRD42020203998).

Results: Eleven studies met the inclusion criteria. The relationship between 1 or more UI features and 1 or more DQ indicators was examined. UI features were classified into 4 categories: 3 types of data capture aids, and other methods of DQ assessment at the UI. The Weiskopf et al measures were used to assess DQ: completeness ($n = 10$), correctness ($n = 10$), and currency ($n = 3$). UI features such as mandatory fields, templates, and contextual autocomplete improved completeness or correctness or both. Measures of currency were scarce.

Discussion: The paucity of studies on UI features and DQ underscored the limited knowledge in this important area. The UI features examined had both positive and negative effects on DQ. Standardization of data entry and further development of automated algorithmic aids, including adaptive UIs, have great promise for improving DQ. Further research is essential to ensure data captured in our electronic systems are high quality and valid for use in clinical decision-making and other secondary analyses.

Key words: electronic health records; data quality; user interface; clinical decision support; standardized data.

Background and significance

Electronic health records (EHRs) include information about patients' health status, care decisions, care received, and care underway. Such information is often used by clinicians (nurses, doctors, and other health professionals) to support decision making in day-to-day care. When aggregated through secondary analyses, it can serve as the foundation for delivering robust clinical decision support (CDS) in the form of alerts and recommendations to clinicians at the point of care.[1] High-quality CDS has been shown to increase the effectiveness of clinical decision-making and reduce documentation time.[2] For EHR data to be genuinely useful to clinicians in day to day practice and as meaningful CDS, however, the data collected (eg, assessments, medication administration,[3] progress notes,[4] and nursing care plans[5]) must be of high quality. Though there are no standard definitions, many experts agree that the main dimensions of data quality (DQ) are completeness, correctness, and currency.[6–10] Data are: (1) *complete* if the expected patient data are present;[7,8] (2) *accurate or correct* if the data are true, truly reflective of real observations, in agreement with other relevant elements or coherent with what the element is measuring;[7,8] and (3) *current* if the data were recorded in the time frame of interest.[6,7,9,10]

The design and features of the EHR user interface (UI), the computer screens with which the user interacts,[11–13] have a potential impact on the quality of data gathered in the EHRs.[12,14,15] A well-designed UI should facilitate the generation of high-quality EHR data by providing the content, features, and functionality that support clinicians to accurately document desired data.[9,14,16] The requisite components of the UI needed to produce high-quality data vary based on the

Received: March 30, 2023; **Revised:** August 22, 2023; **Editorial Decision:** August 30, 2023; **Accepted:** September 5, 2023

purpose and type of data being documented (eg, care planning information, vital signs, discharge status) and the structure of the data (eg, structured, unstructured data, or both).[14]

Recently, concerns have been raised about the unintended consequences of poorly designed EHR UIs.[11,17–19] Though the application of usability metrics has become popular to limit interface design flaws, improve the user experience, and reduce errors, these metrics do not routinely include explicit measures assessing the impact on DQ.[20–22] Poorly designed UIs can cause data entry errors, failure to capture needed information, and dissemination of inaccurate information potentially resulting in dire consequences for patients.[12,23] Further, use of these inaccurate data in CDS algorithms will result in generating invalid CDS[3,15,24,25] that may go undetected. Metrics continue to evolve for evaluating the DQ at the backend (database side) but these do not adequately address the errors that can occur at data entry or UI side.[7] The importance of understanding the relationship between the EHR UI data collection features and the quality of data captured in the EHR database is just now gaining attention.[11] Since the data collected in EHRs are used in care decisions of all types that directly influence patient outcomes, it is crucial to determine what is known about the relationship between UI features and the quality of data collected. Although a positive usability assessment is evidence of user acceptance of a UI, it does not fully address the relationship between the UI and DQ. Below we report the results of our comprehensive integrated literature review to address this gap.

Objective

The specific aim of this integrative review is to synthesize findings from primary research studies to better understand what is known about the relationship between EHR UI features and the quality of data.

Materials and methods

This was an integrative review following the updated methodology proposed by Whittemore and Knafl.[26] The identification of the literature; extraction of the concepts, patterns, and relationships between UI features and DQ; and synthesis involved multiple iterations (PROSPERO identification number: CRD42020203998).[27]

Literature sources and search strategy

Our reproducible search strategy was developed with the assistance of a health sciences librarian (M.A.) with no set begin date and an end date of October 10, 2022. We searched 5 databases (EMBASE, CINAHL, Web of Science, PubMed, and APA PsycInfo) using a common strategy for all but adapting it to the nuances of each database. The search strings thus varied by database, but all included the key concepts of "user interface," "electronic health records," and "data quality" connected with logical operators "and" and their equivalent terms connected with "or." We downloaded the records from the search results into an electronic reference manager (EndNote version X9) and systematic review production tool (Covidence). Appendix S1 displays the details of the database search strategies.

Inclusion and exclusion criteria

Our inclusion criteria for selecting eligible studies were: (1) peer-reviewed, (2) written in English, (3) focused on the UI data entry features and functionality of systems (eg, EHRs) that captured health care information, (4) included clinicians (eg, nurses, physicians, and other health providers), and (5) included at least 1 measure of DQ. Exclusion criteria were: (1) opinion articles, case studies, conference abstracts, letters, newspapers, and editorials; (2) studies without clinicians; (3) studies that focused on data entry with no DQ component, and (4) studies that examined DQ with no data entry component. We did not exclude articles based on dates or study design.

Screening and selection

Following the literature search, all identified articles were downloaded and duplicates removed. The title, abstract screening, and full-text reviews were performed independently by individual reviewers in 2 different dyads; (O.O.M.) and (F.D.S.) reviewed nearly 95% of the articles with (O.O.M.) and (G.M.K.) reviewing those remaining. The cumulative interrater reliability was 95.2% with a Cohen Kappa of 0.45 indicating moderate agreement. Disagreements were settled through discussion during the abstract and title review phase. When consensus could not be reached, the article was moved to full-text review. The full-text review process was conducted similarly by both dyads with disagreements being settled through consensus and when needed by a third reviewer (R.I.B.).

Extraction of review article information

For each of the included articles, we extracted: (1) reference information, (2) purpose of study, (3) description of the UI, (4) study design, sample, data source, and study location, (5) measures of DQ and usability (if assessed), (6) results, and (7) the study quality assessment rating. The extraction was iterative and validated by 4 co-authors (O.O.M., R.I.B., G.M.K., and Y.Y.) to ensure the desired content was captured clearly and accurately. Based on the content extracted, we assessed the quality of each study article.

The DQ measures identified in each review article were classified into categories identified by Weiskopf et al,[8] 3 of which are identified as core by most experts: (1) *completeness*: if the expected patient data are present;[8] (2) *correctness*: if the data are accurate, truly reflective of real observations, in agreement with other relevant elements or coherent with what the element is measuring;[8] and (3) *currency*: if the data were recorded in the time frame of interest.[7,8] All studies included a UI description (an inclusion criteria).

Appraisal of the quality of studies

We adapted Kmet, Lee and Cook quality assessment tool[28] to evaluate the quality of the studies (see Table 1). The updated version of the risk bias quality assessment tool helped us focus on rating the strength of the elements of the studies of interest to us. The 4 areas examined included: (1) system description (1 item), (2) usability assessment (3 items), (3) DQ assessment (3 items), and (4) study design and findings (3 items). Two raters (O.O.M. and R.I.B.) used the quality assessment tool to independently evaluate the studies. Raters reached a consensus after discussing their differences in scores, and the third rater (G.M.K.) reconciled disagreements. Each of the 10 items could be scored on a 0-2 scale with a possible total score range of 0-20, with higher scores indicating higher quality of studies for the purposes of this review.

Table 1. Risk bias quality assessment tool—updated version.

Tailored components	2	1	0
Intervention/technical system			
1. Completeness of user interface design description. (*Interface design=description of the features a subject interacts with in electronic system supporting data entry*)	High	Moderate	No description
Usability			
2. Extent to which usability of the interface was evaluated[a] (*based on number of standard usability components evaluated)*	>4	2-4	0-1
3. Quality of assessment tool used to evaluate user interface.	Validated tool or reference to an existing measure assessment	Nonvalidated tool and no reference to an existing measure assessment	No tool or measure assessment
4. Rigor of sampling method/s involving users.	Random selection	Nonrandom selection	No sample
Data quality			
5. Extent to which data quality were included[b] (*based on standard data quality measures*)	3+	1-2	0
6. Quality of assessment tool used to evaluate data quality	Validated tool or reference to an existing measure assessment	Nonvalidated tool and no reference to an existing measure assessment	No tool or measure assessment
7. Rigor of data selection/ extraction	Random selection	Nonrandom selection	None
Study design and findings			
8. Study included control group or pre-post measures	Control group	Pre-Post	None
9. Rigor and clarity of data analyses	High	Moderate	Low
10. Generalizability of results	2 or more study sites, random selection (of EHR), and large sample size	Small sample size or 1 study site	Qualitative data
Total score			

[a] Usability evaluation, eg, time to document, ease of use, user satisfaction.
[b] Data Quality (DQ) measures: completeness, correctness, currency. Quality score category: Low (0-7), Moderate (8-14), High (15-20). Total quality score per item: Intervention/technical systems (2), usability, Data quality and Study design and findings (6 each).

Results

Our search yielded 1,323 studies, with 30 studies retained to full-text review. Of those, 11 met the inclusion criteria and were included as final studies in the review. The reviewing process and results are depicted in Figure 1.

Study characteristics

The 11 included studies from 9 different countries across 3 continents. Five studies[29–33] came from North America (Canada [$n=1$] and United States [$n=4$]), 4 from Europe[34–37] (Denmark, Sweden, Turkey, United Kingdom, $n=1$ in each), and 2 from Asia[38,39] (China, Israel, $n=1$ in each). Seven of the studies were comparative studies of UIs[29,30,33,36,37,39] or data entry processes,[31] comparing DQ before and after an update implementation. The remaining 4 were cross-sectional studies, examining DQ without a pre-post comparison.[32,34,35,38] Three studies[32,34,35] were reported to have been underpinned by a theoretical model or conceptual framework. One of them followed the Weiskopf & Wang DQ assessment framework,[34] the second applied a usability engineering perspective in the analysis phase,[32] and the third was rooted in narrative theory and distributed cognition.[35]

Eight studies were quantitative,[29,31,33,34,36–39] 2 were qualitative,[32,35] and 1 used mixed methods.[30] The study designs included randomized control trial,[39] experimental design,[33] pre-post comparison,[29,37] retrospective chart review,[31,38] descriptive,[34] within-participant study,[36] mixed methods,[30] qualitative analysis,[32] and ethnographic study.[35] Samples of clinicians were small ranging from 6 to 85 with one of the studies having a sample drawn from more than 1 system or location (sites = 5).[35] Patient record data were included in 10 studies.[29–34,36–39] All but one[34] of the studies that utilized subsets of actual patient record data (cases) extracted these from a single system or utilized hypothetical scenarios.[36,39] In 5 studies,[29,30,37–39] 1 or more aspects of usability (eg, task completion, time required for task, user satisfaction, keystrokes) were also assessed. Usability measures included those locally developed or previously validated instruments such as the System Usability Scale.[22,40] Table 2 displays details of study characteristics and extracted evidence in the Supplementary Material.

Data quality measures

Per our criteria, all 11 studies[29–39] included at least 1 measure of DQ and provided rich descriptions of the UIs and/or screen

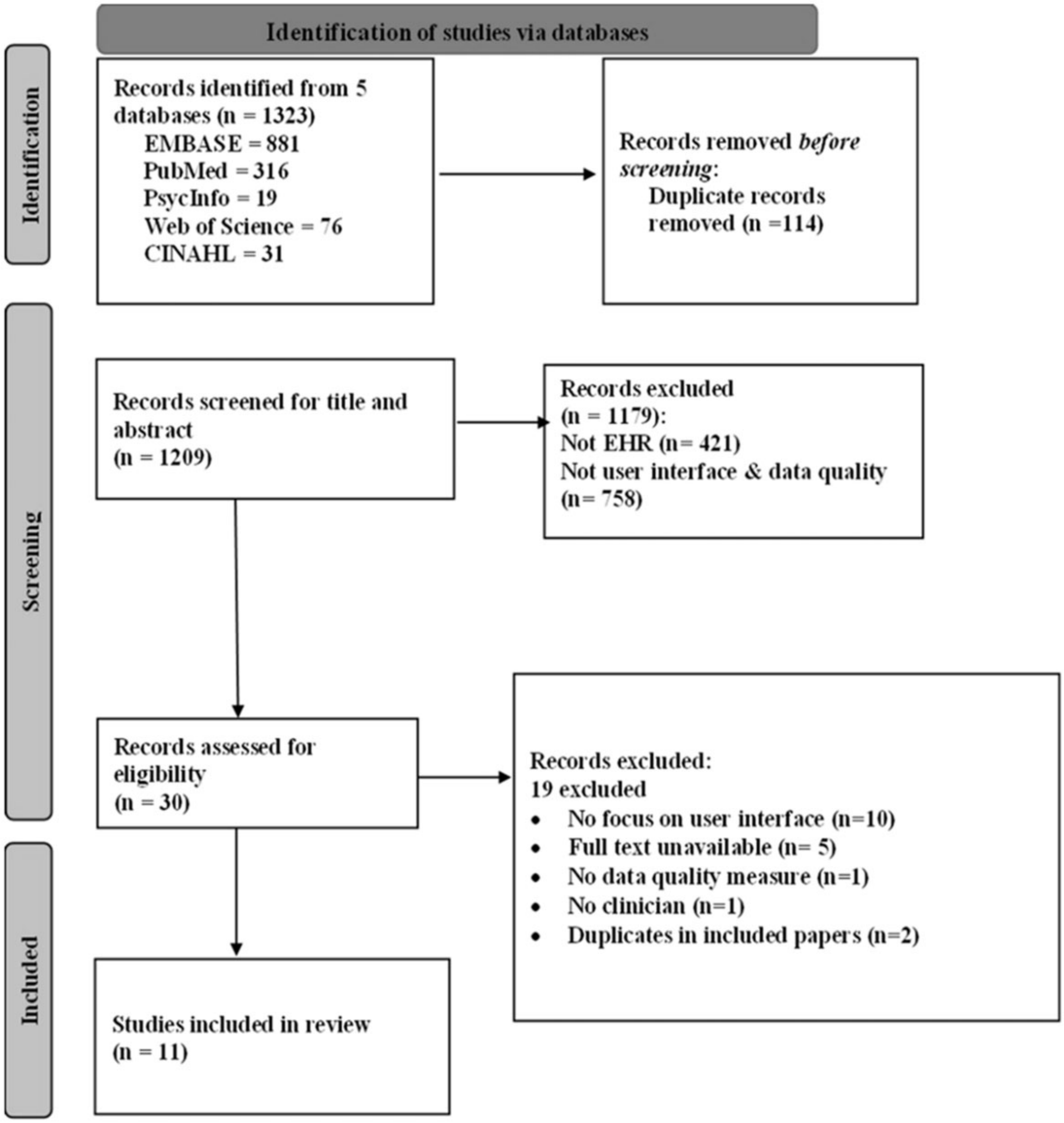

Figure 1. PRISMA flow chart showing search results.

layouts. Nine studies[29–32,34–36,38,39] assessed 2 or more DQ measures and 5 studies reported usability assessments.[29,30,37–39],

Completeness was examined in 10 studies.[29–36,38,39] Across the 10 studies, completeness was analyzed as the percentages or proportions of missing[30,33,38] or complete fields;[29,31,34] counts of complete fields;[39] and/or the mean/average number missing,[33] complete[36] patient information, or perceptions of completeness by nurses[32] or physicians.[35]

Correctness was examined in all but one.[33] Within the construct of correctness, 3 studies explored concordance[34,36,38] (matching information across anesthesia records, concordance across 3 systems, amount of unrelated diagnostic information documented), 4 examined accuracy by comparing expert designed sources (predictive modeling,[39] data modeling,[37] templates,[29,31]), and 3 reported user perceptions of correctness or accuracy.[30,32,35]

Lastly, currency was addressed in 3 studies. One study evaluated currency by assessing the time between entry to emergency room to documentation of vital signs.[34] The 2 qualitative studies provided user perceptions; one described currency contrasting documentation done at the point of care visit versus elsewhere after the visit[32] and the second in terms of chronological coherence.[35]

UI features and DQ

In all 11 studies, at least 1 UI feature was identified and expected to impact 1 or more of the 3 core quality measures. As no *a priori* categorization of UI features linked to DQ was found, we inductively derived categories from our findings in this review and classified these UI features into 1 of 4 categories based on the mode of data capture aid: (1) mandatory data capture aids;[33,38] (2) nonmandatory data capture aids;[29,31] (3) automated algorithmic data capture aids,[30,36,37,39] and (4) other ways to relate UI features and DQ.[32,34,35] Below we report the findings by group category.

Table 2. Extensive evidence reports of 11 studies about the user interface features and data quality.

First author, year	Tailored purpose of study	User interfaces	Study design, sample data source and location	Measures of data quality (DQ) and usability	Results	Quality assessment
Avidan and Weissman[38]	To determine completeness, concordance, and user acceptance of Anesthesia Information Systems (AIMS), a customizable data entry product that replaced an existing system 1-year postimplementation.	**Original interface:** A variety of features streamlined for data entry, such as user response in one field links to applicable follow-up fields (eliminating the need to view nonapplicable fields). Saving was permitted only after all mandatory fields were completed. **New interface:** N/A	**Study design:** (a) retrospective chart review of 4 mandatory forms, (b) satisfaction survey of clinicians **Sample:** 12,287 anesthesia records of one hospital's anesthesia department—entered by 60 Anesthesiologists (in 17 operating rooms). Sub-sample: 5,626 patients with age record and one or more of following: • Tracheal tube (TT), $n = 1,848$ • Laryngoscope blade (LB), $n = 1,967$ • Intravenous catheter (IVC), $n = 3,464$ • IV site insertion • Operating room (OR), $n = 7022$ • Ward, $n = 2710$ User satisfaction: 49 surveys were completed by anesthesiologists. **Theories/models:** None **Data source:** Anesthesia data **Location:** Israel	**Completeness:** % anesthesia records captured in 4 mandatory forms. **Concordance:** Distribution of TT, LB and IVC sizes versus age groups (0-2, 2-5, 5-10, 10-18, >18); relationship between the variables. IV insertion site: % frequency of IV insertion on left arm in OR and comparison with % of left arm insertion ward **User satisfaction and usability:** Clinician's acceptance and usability of AIMS. **Analytical approach:** Counts, percentages; chi-square test and Pearson's correlation coefficient	**Completeness:** 99.6% of all anesthesia records were complete. **Concordance:** Significant association between TT size and age ($P < .0001$), between LB size and age ($P < .0001$), and between IVC size and age ($P < .0001$). IV insertion site: 73.8% of left arm IV insertion in OR and 52.9% in ward; IV insertion on left arm significantly higher in OR ($P < .0001$). **User satisfaction and usability:** Proficiency: 80% good-very good Usability: 88% good-very good Form design: 86% good-very good	Moderate
Adams et al29	To determine change in completeness, user satisfaction, and documentation time following the introduction of an immunization entry software for children's immunization visits.	**Original interface/Process:** Complex multistep human process that required data entry on paper. **New interface:** Immunization tracking software upgraded to have predefined standardized list of vaccinations that can be selected using the keyboard, eliminating the need to type or handwrite information.	**Study design:** Pre- and postimplementation of new electronic tracking system **Sample:** 1 primary care center with 488 children (<5 y/o) visits: *231 pre; 257 post* User Satisfaction/Documentation time: Survey administered to 9 nurses. **Theories/model:** None **Data source:** Vaccine data **Location:** United States	**Completeness:** % Immunization visits and doses missing in database. **Correctness:** % of inaccurate on-schedule data in database. **User satisfaction:** Nurses' satisfaction 3-month postimplementation. **Documentation time:** Percentage of nurses who deemed the system as time saving. **Analytical approach:** Descriptive analysis; Percentages	**Completeness:** Dose missing: 37.9% (pre) versus 0% (post) **Correctness:** Incorrect on-schedule data: 31.6% (pre) versus 8.6% (post) **User satisfaction:** 100% recommend regular use of software **Documentation time:** 89% agreed that the system saved time.	Moderate
Greenbaum et al30	To determine the effect of a domain-specific ontology and machine learning-driven user interfaces on the efficiency, completeness, accuracy of documentation of presenting	**Original interface:** Unstructured free text entry of presenting problems. **New interface:** (1) Unstructured free text entry of presenting problems and	**Study design:** Mixed methods study; pre-and postintervention design, and a qualitative study **Sample:** 1 hospital trauma center. • Pre = 55 nurses; 56,186 patient visits	**Completeness:** Structured data capture (% patients whose presenting problems can be automatically mapped to HaPPy) Reviewer ratings of completeness (rating 1-4).	**Completeness:** Structured data capture: 26.2% (pre) vs 97.2% (post), $P < .0001$. Completeness rating: 3.35 (pre) vs 3.66 (post), $P = .0004$	Moderate

(continued)

Table 2. (continued)

First author, year	Tailored purpose of study	User interfaces	Study design, sample data source and location	Measures of data quality (DQ) and usability	Results	Quality assessment
	problems in the emergency department (ED).	(2) Structured data entry using the top 5 suggested presenting problems (clicking) and contextual autocomplete (typing the first few letter) that ranked concepts by the predicted probability via machine learning coded in HierArchical Presenting Problem ontologY (HaPPy) from triage vital signs, demographics, and a brief triage note.	• Development = 85 nurses; 144,077 visits • Post = 53 nurses, 77,303 patient visits 3 reviewers independently rated 150 random patient records (50 pre, 50 post, 50 during unscheduled system downtime) on completeness, precision, and overall quality using a 4-point Likert Scale. **Theories/model:** None **Data source:** Medical-surgical data **Location:** United States	**Accuracy:** Reviewer ratings of precision and quality (rating 1-4). **Keystrokes:** Number of keystrokes per presenting problem. **Analytical approach:** Support vector machine; Fischer's exact test and *t*-test.	**Accuracy** Precision rating: 3.59 (pre) vs 3.74 (post), $P=.0998$. Quality rating: 3.38 (pre) vs 3.72 (post), $P=.0002$. **Keystrokes:** 11.6 (pre) vs 0.6 (post), $P<.0001$.	
Skyttberg et al[34]	To compare 3 documentation platforms in emergency units on vital sign (v/s) data quality.	**Original interface:** 1) paper-based documentation, 2) mixed documentation (paper data entry manually transferred into EHR), 3) direct documentation into EHR **New interface:** N/A	**Study design:** Descriptive study **Sample:** v/s data from 335,027 emergency care visits of patients >18y/o in 2013 at 5 hospitals 1) Paper-based: Hospital 1 = 59,679 Hospital 2 = 62,764 2) Mixed: Hospital 3 = 59,900 3) Electronic: Hospital 4 = 78,991 Hospital 5 = 73,693 **Theories/model:** Weiskopf & Wang data quality assessment framework **Data source:** Medical-Surgical data **Location:** Sweden	**Completeness:** %v/s in EHR **Correctness/accuracy:** Plausibility and concordance as a proxy **Currency:** timeliness of v/s documentation into EHR (relative to patient arrival) **Analytical approach:** Descriptive analysis.	**Completeness:** 1) Paper-based • Hospital 1 = 2% • Hospital 2 = 1% 2) Mixed • Hospital 3 = 95% 3) Electronic • Hospital 4 = 71% • Hospital 5 = 62% **Correctness/accuracy:** High for all 3 systems. **Currency:** 1) Paper-based: Low 2) Mixed: Low 3) Electronic: Medium	Low
Urchek et al[31]	To compare completeness and accuracy of templated versus nontemplated documentation tool for entering clinical examinations of pediatric supracondylar fractures.	**Original interface:** Nontemplated form for documentation. **New interface:** Templated form for structured entry of clinical examination notes.	**Study design:** Retrospective Chart Review **Sample:** 119 children (1-15 y/o; 42 templated vs 77 nontemplated) with supracondylar fractures at 1 pediatric hospital **Theories/model:** None **Data source:** Orthopedic data **Location:** United States	**Completeness:** % Patients with complete history and physical (H&P) documentation. **Accuracy:** % Accurate documentation of nerve palsies (comparison with attending's notes). **Analytical approach:** Percentages, chi-square test of homogeneity	**Completeness:** 67% (templated) vs 10% (nontemplated), $P<.001$. **Accuracy:** 84% (templated) vs 93% (nontemplated), $P=.16$	Moderate
Yang et al[32]	To examine how the goals of home care nurses affect their EHR documentation strategies.	**Original interface:** Laptop-based commercial EHR with varying interface controls.	**Study design:** Qualitative study **Sample:** A convenience sample of 1 rural home health agency.	**Completeness:** Reason and action taken to complete entry at patient home or a later time.	**Completeness:** Nurses start documentation at the patient's home and have to complete/modify the entry later because EHR entry screen	Low

(continued)

Table 2. (continued)

First author, year	Tailored purpose of study	User interfaces	Study design, sample data source and location	Measures of data quality (DQ) and usability	Results	Quality assessment
		New interface: N/A	10 home care patients. 5 white female registered nurses, 25-47 y/o. **Theory/model:** Usability engineering perspective for analysis **Data source:** Care plan data **Location:** Pennsylvania, United States	**Accuracy:** Perceptions of correct information based on goals impacting documentation strategies. **Currency:** Reduced time: reason for completing documentation at a later time. **Analytical approach:** Thematic analysis	allows it. **Accuracy:** 1 nurse was concerned because they sometimes click wrong button. 1 saves incomplete text in entry screen for later completion. **Currency:** To reduce time, 2 nurses completed their data entry later to reduce time spent at patient visit and to avoid infection.	
Jensen and Bossen[35]	To examine the use and reasons for use and nonuse of the overview interface in EHR.	**Original interface:** An overview screen display of patient information summarizing integrated patient information from diverse sources in the EHR. **New interface:** N/A	**Study design:** Ethnographic studies **Sample:** 380 departments from 5 hospitals Snowball sample of 6 Physicians (3 users and 3 nonusers) from 5 different departments by information technology (IT) staff. 8 physicians (5 users and 3 nonusers) from 4 departments with most uses (2 anesthesiologic and 2 outpatient) by head of departments **Theories/model:** Narrative theory and distribution cognition **Data source:** Clinicians and IT staff data **Location:** Denmark	**Completeness:** Perception of the completeness for use/nonuse of patient summary data. **Accuracy:** Perception on the trustworthiness for use/nonuse of patient summary data. **Currency:** Perceptions on the chronological coherence of data No direct measure of DQ, but addressed the following: **Use:** # interface uses per day for a department **Analytical approach:** Thematic analysis; counts, and percentages	**Completeness:** Lack of trust of completeness was a reason for nonuse of patient summary data. **Accuracy:** Lack of trust in accuracy of information was a reason for nonuse. **Currency:** Concerns about the chronological coherence of data. **Reason for use** Clinical work largely standardized, less distributed, less complex. **Use:** $\leq$1/day (65%) 2-3/day (14%) $\geq$3/day (21%)	Low
Hua and Gong[39]	To explore text prediction techniques that enhance quality and efficient data entry	**Original interface:** Two interfaces: <u>Control:</u> Structured data entry via multiple choice questions and unstructured data via free text and comment fields with no text prediction capability. **New interface:** <u>Treatment—</u> Cueing list as added feature to structured data entry via multiple choice questions and unstructured data via free text and cueing list and auto-suggestion as added features to comment fields.	**Study design:** Randomized control trial **Sample:** <u>52 nurses</u> Control: 25 nurses Treatment: 27 nurses 260 reports of patients' falls for 5 case scenarios Usability: 52 surveys (5-point rating) **Theory/model:** None **Data source:** Patients falls data **Location:** China	**Completeness:** Counts of verified narrative text and missing information in comments field. **Accuracy:** Percentage of structured entry accuracy **Usability:** Nielsen's Attitudes of Usability—measured learnability, efficiency, memory, satisfaction **Keystrokes:** Number of mouse clicks and keystrokes **Analytical approach:** Descriptive and regression analysis	**Completeness:** <u>Control vs Treatment group</u> Narrative (free text): 3.8 ± 2.3 vs 5.1 ± 2.4, $P = .000$ Comments field: 20/125(16.0%) vs 2/135(1.5%), $P = .000$ **Accuracy:** <u>Control vs Treatment group</u> 79.4 ± 10.1% vs 83.2 ± 11.0%, $P = .000$ **Usability** <u>Control vs Treatment group</u> Learnability 4.0 ± 0.3 vs 4.1 ± 0.5, $P = .545$ Efficiency 4.1 ± 0.4 vs 4.3 ± 0.6, $P = .386$ Memory 3.5 ± 0.5 vs 3.8 ± 0.5, $P = .099$ Satisfaction 3.8 ± 0.4 vs 3.9 ± 0.5, $P = .458$ Keystroke 173.2 ± 117.0 vs 89.48 ± 9.4, $P = .000$ Mouse clicks 13.3 ± 2.1 vs 17.2 ± 3.7, $P = .000$	High

(continued)

Table 2. (continued)

First author, year	Tailored purpose of study	User interfaces	Study design, sample data source and location	Measures of data quality (DQ) and usability	Results	Quality assessment
Kaka et al33	To compare data collection platforms (paper and electronic) developed to improve data entry and data quality	**Original interface:** Paper **New interface: electronic:** Forms with entry field that requires saving data before navigating to next page and highlights missing fields before submitting forms	**Study design:** Experimental study (random assignment) **Sample:** 106 patients (53 each for paper and electronic) Physicians Paper: 44 Electronic: 48 **Theory/model:** None **Data source:** Psoriasis and psoriatic arthritis patients **Location:** Canada	**Completeness:** Measured mean of incomplete physician entries of all patient data. **Documentation:** Average time taken to enter data and patient-physician encounter time. **Analytical approach:** Descriptive analysis: mean (SD), and *t*-test for comparison of the means of 2 samples.	**Completeness:** Paper vs Electronic 3.34 ± 3.24 vs 2.19 ± 2.32 **Documentation time:** Data entry time: 14.2 ± 7.69 vs 16.14 ± 6.48, $P = .20$ Patient-physician encounter (paper vs electronic): 46.5 ± 15.9 vs 37.2 ± 10.5, $P = .0035$	Moderate
Kostopoulou et al36	To compare the data completeness and correctness in consultations entry software with and without a clinical decision support system (CDSS).	**Original interface:** Text search feature with no predictive capability for finding clinical data elements. **New interface:** Ontology-driven predictive text capabilities and drop-down menu for suggested list of symptoms associated with diagnoses.	**Study design:** Within-participant study **Sample:** 34 general practitioners 12 patients With CDSS—6 patients Without CDSS—6 patients Hypothesis 1: More data items to be documented system with CDSS than one without CDSS. Hypothesis 2: Entry software with CDSS contains a lower proportion of diagnosis-related data items than system without CDSS. **Theory/model:** None **Data source:** Clinical consultation data **Location:** United Kingdom	**Completeness:** Measured the means of documentation captured and incidence rate ratio between the entry software with CDSS and one without CDSS. **Correctness:** Measured the means of unrelated (biased) entries and proportion of diagnosis-related data in entry software with CDSS and one without CDSS. **Analytical approach:** Descriptive analysis; Regression analysis; Wilcoxon matched-pairs signed-ranks; and sign tests.	**Completeness:** Without CDSS All data (codes and free text)—12.15 ± 4.00 Codes—2.15 ± 2.53 Free text—10.00 ± 4.24 Diagnosis-related data—8.40 ± 2.86 With CDSS All data (codes and free text)—15.73 ± 5.22 Codes—12.39 ± 5.33 Free text—3.34 ± 3.03 Diagnosis-related data—9.60 ± 3.36 Hypothesis 1 Overall—1.29 [1.18, 1.42] $P < .001$. Coded—5.76 [4.31, 7.70], $P < .001$ Free text—0.32 [0.27, 0.40], $P < .001$ **Correctness:** Without CDSS All data (codes and free text)—0.71 ± 0.17 Codes—0.78 ± 0.33 Free text—0.72 ± 0.21 With CDSS All data (codes and free text)—0.63 ± 0.17 Codes—0.65 ± 0.19 Free text—0.54 ± 0.37 Hypothesis 2 Overall: −0.08 [−0.11 to −0.05], $P<.001$). Codes: −0.13 [−0.19 to −0.08], $P < .001$. Free text: −0.18 [−0.24 to −0.12] $P < .001$	Moderate

(continued)

Table 2. (continued)

First author, year	Tailored purpose of study	User interfaces	Study design, sample data source and location	Measures of data quality (DQ) and usability	Results	Quality assessment
Kuru et al37	To validate the quality data collection in SISDS method, based on user acceptability and performance of SISDS compared to the existing approaches.	**Original interface:** Transcriptionist oriented approach **New interface:** Free text data entry field designed with inline editing for making changes to contents on entry screens while user is still entering data without toggling between edit and read view.	**Study design**: Descriptive study design and pretest and posttest design **Sample:** Transcriptionist oriented approach versus SISDS Methods 8 radiologists examined each method 16 cases 3 significant cases diagnosed by ICD codes K21.9 = 250 instances K44.9 = 126 instances K22.4 = 116 instances Acceptance and usability 16 criteria survey questions to 20 physicians and clinicians **Theory/model:** None **Data source:** Esophagus data **Location:** Turkey	**Accuracy:** % difference of correctly diagnosed cases by radiologists in SISDS method and transcriptionist approach **Acceptability:** Measured radiologists and clinicians' satisfaction from survey **Analytical approach:** Descriptive analysis	**Accuracy** Transcriptionist oriented approach 7 cases of K22.4: correctly diagnosed. Success rate = 43.75% (7/16) SISDS Method 13 cases of K22.4 correctly diagnosed Success rate = 81.25% (13/16) **Acceptability:** SISDS rated 25 out of 32; high satisfaction compared to other reporting approach	Moderate

Mandatory data capture aids

This feature requires users to complete data fields before being allowed to move to the next task. Two studies reported the use of mandatory fields. In 1,[38] the feature was used to capture important anesthesia details. Completeness after 12 months was at 99.6% with correctness measured through concordance with significant associations ($P<.0001$) between age and tracheal tube, laryngoscope, and catheter sizes. In the second study, as an alternative to paper entry, clinicians were provided an electronic tablet to enter key information on patients with psoriatic arthritis that disallowed submission until all highlighted fields were completed. The authors reported improvement in completeness while increasing documentation time.

Nonmandatory data capture aids (checklists and templates)

This category included 2 studies comparing free text data capture to data entered using a checklist[29] or template.[31] Checklists and templates of data items are typically evidence based and set-up as memory aids to remind users to adequately characterize, deliver, and document appropriate care. In 1 study,[29] researchers sought solutions to reduce missing vaccine doses in a multistep immunization data entry process.[29] By switching to a single-point data entry system with selections from predefined standardized vaccine information lists, they improved the completeness and correctness by reducing missing doses from 37.3% to 0%.[29]

A second study addressed completeness of pediatric orthopedics data by replacing nontemplate (unstructured) forms with template forms for structured documentation of clinical examination.[31] After implementing template forms, the completeness of patient history and physical documentation rose from 10% to 67% ($P<.001$). No significant difference was found between the proportions of accurate documentation of nerve palsies (assessed through comparison with attending's notes) in the template form (84%) versus the nontemplate forms (93%) ($P=.16$).[31]

Automated algorithmic data capture aids

There were 4 studies in this category representing different purposes and varying degrees of complexity in the type of model or algorithms used to support DQ. Three studies focused on capturing structured data[30,36,39] with one[39] also including unstructured data capture. One used a nationally recognized standardized terminology by mapping data to a domain-specific ontology.[30] In these studies, predictive algorithms that generated auto-suggestions were used to enhance the completeness and accuracy of patient data entered: for falls,[39] presenting problems,[30] and symptoms and diagnoses.[36] Completeness was significantly higher in the Hua and Gong[39] study on unstructured falls data (34%, $P=.000$) in the treatment (auto-suggestions) versus control group (no auto-suggestions). Significant differences in completeness of structured data capture were found for the treatment groups who received auto-suggestions for problems in the Greenbaum et al[30] study ($P<.0001$) and for problems and symptoms in the Kostopoulou et al[36] study ($P<.001$) versus the groups receiving none. Correctness was also examined in all 3 studies.[30,36,39] In the falls study, the treatment group (cueing lists) had significantly higher correct structured data capture (3.8%, $P=.000$)[39] compared to those receiving no cueing lists. In the study on symptoms and diagnosis, the treatment group had significantly higher correctness ($P<.001$)[36] than the control group measured as amount of unrelated diagnoses data entry. In the presenting problems study no significant differences, as measured by user perceptions, were found between the treatment and control groups.[30]

In the fourth study, Kuru et al[37] examined the reporting of esophagus related data by radiologist. For the treatment group, in-line editing based on the Structured, Interactive, Standardized, and Decision Supporting Method (SISDS) developed model and algorithm, was provided to enhance the correctness of data captured.[37] A success rate of 81.25% in capture of correct diagnosis was achieved with the SISDS group compared to 43.75% for the transcriptionist control group.[37]

Other ways to relate UI features and DQ

There were 3 studies in this category, all examining completeness, correctness, and currency.[32,34,35] In 1 study, 3 vital sign data entry formats were compared across 5 institutions.[34] The other 2 are qualitative studies; one assessed DQ of first home visit documentation in 1 EHR;[32] and the second examined physician perceptions of DQ of content in an EHR overlay screen.[35]

In the first study, Skyttberg et al[34] compared DQ of 3 different formats; paper-based, mixed (paper transferred to electronic), and electronic of vital signs of data entry from 5 emergency rooms. The percentage of vital signs captured (completeness) was highest in the mixed (M) documentation format (95%), followed by electronic (E) data entry (62%-71%), and finally paper-based (P) entry (1%-2%).[34] Correctness was measured as concordance with vital sign norms and found to be high across all 3 formats. Currency was measured as the time between admission to EHR and entry of vital signs into the record (with "15" was the desired benchmark).[34] Paper-based and mixed had low currency and the electronic format had medium currency.[34]

The 2 qualitative studies in the review focused on nurses[32] and physicians[35] perceptions about UI features that influence DQ. In the Yang et al[32] study, home care nurses from 1 agency were queried about their reasons for delaying full documentation of a patient's first home visit until a later time and place away from the patient. The nurses indicated that documentation requirements and system design prevented complete data entry at the point of care (compromised currency). Work arounds were created and used as memory aids (eg, entering some key data into EHR and creating written notes) for the purpose of ensuring completeness and correctness at the time of delayed data entry.[32]

In the second qualitative study Jensen et al,[35] conducted ethnographic research with a small sample of physicians from 5 hospitals to examine reasons for use and nonuse of the UI overlay screen. The overlay was originally designed to provide a comprehensive view of patient's key data points extracted from various parts of the EHR. The study was motivated by learning that there were a surprisingly low percentage of users. The findings indicated the need to consider the complexity of the patient population and physician specialty when designing UI features intended to deliver high-quality data. Some found the overlay sufficiently complete, correct, and current. Other physicians found the data incomplete, inaccurate, and lacking in chronological coherence resulting in the lack of trust and nonuse of the overlay.[35]

Table 3. Quality assessment result of 11 studies from the updated risk bias quality assessment tool.

First author	Intervention/technical systems	Usability			Data quality			Study design and findings			Quality score[a] ($n = 20$)
	I.1	I.2	I.3	I.4	I.5	I.6	I.7	I.8	I.9	I.10	
Hua and Gong39	2	2	2	2	1	1	2	2	1	1	16
Avidan and Weissman38	2	1	1	1	1	1	1	0	2	1	11
Adam et al29	2	1	1	1	0	1	1	1	2	1	11
Greenbaum et al30	2	0	0	1	1	1	2	1	1	1	10
Kostopoulou et al36	2	0	0	0	1	1	1	2	1	1	9
Kuru et al37	2	0	1	1	1	1	1	0	1	1	9
Kaka et al33	2	0	0	0	1	1	2	0	1	1	8
Urchek et al31	2	0	0	0	1	1	1	0	2	1	8
Skyttberg et al34	1	0	0	0	2	1	1	0	0	1	6
Yang et al32	2	0	0	0	1	1	1	0	1	0	6
Jensen and Bossen35	1	0	0	0	1	1	1	0	0	0	4

[a] Total score (n) = 20; Quality score category: Low (0-7), Moderate (8-14), High (15-20). Note: Studies were scored based on the updated version of the Risk Bias Quality Assessment Tool. I.1: Completeness of user interface design. I.2: usability evaluation. I.3: Quality of the usability tool used to evaluate user interface. I.4: Rigor of sampling method/s involving users. I.5: Extent to which data quality was evaluated. I.6: Quality assessment tool used to evaluate data quality. I.7: Rigor of data selection/extraction. I.8: Study included control group or pre-post measures. I.9: Rigor and clarity of data analyses. I.10: Generalizability of results.

Quality appraisal of studies

Quality Scores ranged from 4 to 16 (total possible = 20), with one of the studies ranking at high quality,[39] 7 at moderate[29–31,33,36–38] quality and 3 studies at low[32,34,35] quality (Table 3).

Discussion

Given the growing emphasis on secondary uses of EHR data (eg, primary data compiled into multidimensional reports, CDS, research etc.) and availability of powerful analytical tools and methods (eg, artificial intelligence), it is crucial to ensure high quality of the data entered into the EHR. As noted, this integrative review of 11 studies represented a wide range of study designs, rigor, UI features, data types involved, and measures of DQ. The designs of the 11 studies varied from descriptive qualitative studies to 1 RCT with the majority involving comparative designs. Clinician sample sizes across the studies were small with the data source being unique to each study. Measures used to assess DQ were diverse, making clear comparisons across studies difficult. Limited sample sizes and varying quality of evidence hinder our ability to make inferences about the relationship between particular UI features and quality of data. To ensure robust findings that are generalizable, future research should emphasize randomized controlled trials with powered samples to investigate the impact of EHR UI features on DQ. There are nonetheless a number of important findings about DQ and the relationship between UI features and DQ that can be drawn from this review.

As noted, the reviewed studies examined DQ measures consistent with the core set of correctness, completeness, and currency identified in the literature.[6,7] However, authors in our review used a variety of measures to assess these categories in order to accommodate the diverse data sources. Several researchers also used indirect proxies as validation when direct observation was not possible. For example, tube size could be validated as correct by a second observer or by use of concordant measures in which the documented tube size is considered valid if it matches the range of sizes recommended for the patient's age.[38] While it makes sense to tailor the measures to the differing data sources, some standardization of measures is necessary to enable comparisons across similar datasets. This requires development or validation of standardized measures with high precision while ensuring that the context of data is adequately captured.

Currency, though recognized as a pivotal measure as indicated in the literature,[6,7] was assessed in only 3 of the 11 studies in this review.[32,34,35] This suggests the potential lack of robust interface (UI) features to ensure data currency. Knowledge of the time or time period covered for the events entered into the record is essential to sequencing events and evaluating impact on outcomes over time. Thus, systems should support its capture. As was reported, there is often a delay in reporting information due to context factors (eg, need to give comprehensive hands on care to multiple patients before charting on them). Factors such as patient workload and acuity also delay timely entry.[41] One potential solution could be to expand use of equipment at the point of care that automatically collects and transfers data and the time of capture to the EHR (eg, vital signs, cardiac rhythms, and intravenous drips).[41] Another is to use voice activated charting which could allow the clinician to document key information into the computer while caring for the patient.[42] While automation might not suit all data types, we should strive to develop approaches that enable instant quality data capture at the point of care. These innovative UI features would also ensure that data-driven clinical decision-making is based on the current condition of the patient at the time of intervention.

Consistent with findings from a recent review on the practice of DQ evaluation, completeness and correctness were commonly assessed.[7] Of the 11 included studies in this review, 10 studies assessed completeness and 10 assessed correctness with 9 evaluating both. As such, completeness and correctness stood out as pivotal DQ measures. UI features such as mandatory fields (forced data entry)[33,38] and predefined templates[31] and picklists[29] (limited set of responses to check) were reported to be helpful in capturing complete and correct data. However, these features frustrated some clinicians; the former increased documentation time,[33] and the latter often did not contain the desired responses.[29,31] The interface overlay, a feature designed to present a valid general summary (completeness, correctness, and currency) of the patient's circumstance to the clinician, was found to be very helpful to some specialties and unacceptable to others.[35] The finding reinforces the need to tailor DQ interface features to the dataset and user characteristics.

The potential benefits of using structured and standardized terminologies at the UI to support data correctness were highlighted in the included studies in this review. For example, consistent use of the same measures, such as vital sign measurements that are recorded in the exact same way, was found to support easy interpretation and comparison of these data within and across organizations.[34] This avoids the need for others who did not enter the data to standardize them for multiple uses through mapping, a process that will naturally compromise correctness. Use of nationally accepted standardized terminologies at the UI also has the potential to enhance correctness of data entered since responses sets for an element are limited to terminologies that have been systematically developed for the purpose. Such use further supports the ability to easily compare data with all others who utilize the same standardized terminologies, a national goal for making health data interoperable.[43]

Auto-suggestion and autocomplete were UI functionalities that emerged as promising for enhancing data completeness and correctness, especially when powered by predictive models and standardized terminologies. Greenbaum et al's study serves as evidence of the power of combining predictive algorithms with nationally recognized terminologies. Thus, the development and adoption of standardized data elements, especially when combined with intuitive UI features, promise significant enhancements in DQ measures and set the stage for more rigorous data-driven methodologies in future research.

It is important to underscore again, however, that UI features and measurements of correctness and completeness will depend on the data sources and the complexity of them. For instance, in Kostopoulou et al[36] study, the researchers added fields for capturing details of the patient's situation to eliminate errors that occurred with an earlier version of an automated diagnoses selection aid. As a result, the team now uses the amount of other details as a proxy for assessing correctness of diagnosis. Another example is correctness being assessed by user perceptions. An applicable example is clinician perceptions about the correctness of their fellow clinician's data entry captured in the EHR.

In general, we believe that automated algorithmic data capture features that utilize standardized terminologies show the most potential for improving data correctness[30,36,37,39] and completeness.[30,36,39] This is because these features reduce the need for recall while using data entered to predict and deliver most likely responses for the data fields. Automated algorithmic aids including contextual autocomplete and auto-suggestions are typically driven by predictive models that enhance capture of clinical information within a single context of use (eg, entry of a standardized patient chief complaint, enhancing data entry of fall patients). These features have commonalities with the robust adaptive UIs (AUIs) that are gradually gaining traction in healthcare systems.[44–47] AUIs are interactive systems designed to preserve usability of UIs across various contexts of use (user, platforms, and environment) based on adaptation rules.[11,48] Similar to the automated algorithmic aids they also utilize algorithms and datasets to enhance UI features for data entry. AUI features are more sophisticated than those found in our studies because they carefully consider variability in context of use. These interfaces are expensive to develop because they include multiple parameters such as cognitive complexity, behavioral predictability, predictive accuracy, users' autonomy to opt out of adaptation rule, users environment, and interface elements that accommodate both novice and experts.[11,48,49] Their UI features focus on ensuring a high degree of data correctness, speed of data capture.[48,49] Thus, AUI could enhance DQ in healthcare, especially when augmented with generative artificial intelligence (AI) models such as Large Language Models to analyze complex heterogeneous data, generate auto-suggestions that could influence complete, correct, and current data. Further research is needed to explore the relationship between AUIs and DQ.

In summary, the studies collectively provide preliminary evidence of the crucial impact UI features can have on DQ. There are features, such as automated algorithmic capture aids, that have the potential to enhance completeness, and correctness. However, little is known about their impact on currency. This underdeveloped area is critical for clinical decision making, quality improvement, and research. Ultimately, further research is required to understand impact of UI features on DQ.

Limitations of the study

We identified several limitations of this review. First, we might have missed identifying important research studies in the 5 databases searched due to the unique features involved in completing advanced searches (eg, variations in registered terminologies such as Medical Subject Headings [MESH] and keywords). The permutation of keywords to create search strings may have also introduced some bias and resulted in the omission of relevant publications. Finally, we may have failed to identify all of the appropriate terms that characterize this area of science, also causing us to potentially miss applicable published studies.

A second limitation was the use of an untested tool for our risk assessment. When no appropriate tool was found, the Kmet et al[28] tool was modified to help rate study facets pertaining to this review. A third limitation was the difficulty we encountered in synthesizing the study findings due to the wide variations in quality and content of the 11 studies. As a result, we may have inadvertently missed important points. To address this issue, we worked hard to make all tables clean and concise and included Table 4 to simply illustrate important study points.

Conclusion

In summary, we have identified key features and elements ranging from simple to complex automated data capture aids, that support the capture of high-quality EHR data relevant for purposes such as making care decisions, using predictive modeling to generate care suggestions, health care research. However, this area of research is underdeveloped, and gaps exist in the relationship between UI features and DQ. Therefore, we encourage researchers to identify ways the tools and research strategies uncovered in this review can be used to ensure one's EHR dataset of interest (eg, care planning data, diversity inclusion and equity data, emergency room visits, epidemiological data) is capturing complete, correct, and current data. We believe a clear priority for future work is to carefully examine and validate ways UI features can be

Table 4. Summary of studies report on user interface features and data quality.

Datasets and authors	Interface feature/s examined (upgraded feature, compared screens)	Completeness	Correctness	Currency	Results
Studies on mandatory data capture aids					
Anesthesia details such as drugs, patient birthdates, sizes of catheter, blades, tubes. Avidan and Weissman[38]	*AIMS system with context sensitive mandatory data entry field* that enforces complete data and saving of entry before proceeding to next task	X	X		12 months of AIMS use resulted in complete data in 99.6%/12,290 anesthesia records; concordance: significant associations ($P<.0001$) between age and: tracheal tube size, laryngoscopy blades; IV catheters
Psoriasis and psoriatic arthritis patients' data captured by physicians. Kaka et al[33]	*Paper entry* vs *Mandatory entry field* highlighted missing data requiring field completion before submitting	X			Improvement in completeness of psoriatic patient data captured electronically via the mandatory field but increased documentation time
Studies on nonmandatory capture aids					
Immunization record containing doses of vaccines, schedule of vaccine administration. Adam et al[29]	*Complicated human data entry process vs Predefined immunization list* organized for user to select without having to type or handwrite	X	X		Improvement in completion and correctness as measured by a reduction in missed and incorrectly scheduled vaccine doses; documentation time was reduced, and user satisfaction was improved
Documentation of supracondylar fracture in children. Urchek et al[31]	*Unstructured nontemplate forms vs templates* with nonmandatory entry field for structured entry of the children's information	X	X		Improvement in completeness in templates over nontemplates ($P<.001$); difference in correctness (accuracy) not significant ($P=.16$)
Studies on automated algorithmic data capture aids					
Capture of presenting problems with structured terms in information system. Greenbaum et al[30]	*Unstructured text entry vs Using a predictive model and autocomplete to support capture of problems as structured/coded data*	X	X		Improvement in completeness of capture of problems as structured/coded data from 26% problem to 98% ($P<.0001$), perceived: (a) completeness ($P=.0004$), (b) correctness (precision) not significant, (c) decrease in keystrokes ($P<.0001$)
Patients falls report documented by nurses. Hua and Gong[39]	*Structured and unstructured interface without text prediction vs Cueing list and auto-suggestions with predictive capability* to enhance structured and unstructured data entry	X	X		Improvement in treatment over control narrative completeness 34% higher ($P=.0000$) and structured data correctness 3.8% higher (accuracy) ($P=.0000$); difference in usability not significant
Clinical consultation data captured by physicians. Kostopoulou et al[36]	*Text search feature with no predictive capability vs Clinical decision support system (CDSS) with ontology-driven interface with text prediction capabilities and drop-down menu that suggests list of symptoms associated with diagnoses*	X	X		CDSS improved completeness of coded $P<.001$ and less free text $P<.001$). Accuracy significantly improved as measured by increase in unrelated text entries (unbiased) being present $P<.001$)
Patient esophagus-related data reported by radiologists. Kuru et al[37]	*SISDS algorithm for report generation with in-line editing* of free text data that shows revised results without toggling to different screens compared to transcriptionist		X		Improvement in correctness of diagnosis: SISDS 81.25% compared to transcriptionist method 43.75% for the transcriptionist-oriented system and high user satisfaction

(continued)

Table 4. (continued)

Datasets and authors	Interface feature/s examined (upgraded feature, compared screens)	Completeness	Correctness	Currency	Results
Studies on other ways to relate UI features and data quality					
Nurses reasons/content of EHR documentation taking place/not taking place at point of care (POC). Yang et al32	*Documentation platform of one EHR for first home visit*	X	X	X	Nurses engaged in distributed (some in home, some later) documentation due to time constraints, amount required, to avoid infection. At visit partial data entry into EHR and hand notes were created as memory aids to support full documentation later off site. Set-up potentially compromises data completeness, correctness, and currency.
Physicians' reasons for using/not using an overview interface (secondary use of data). Jensen and Bossen35	*Overview user interface*, a summary display field showing summary of patient information	X	X	X	Reasons for using—clinical work largely standardized, less distributed, less complex; Reasons for not using overview; trust issues about source of information (correctness), incomplete data on complex patients, chronological coherence of data (currency)
Vital signs (vs) of patients captured at the emergency unit of 5 different hospitals. Skyttberg et al34	Different data entry platform; *paper-based (PB)*, *mixed (M)* paper with manual transfer to EHR; electronic (E), entered directly into *EHR*	X	X	X	Comparison across 5 sites Low (L), medium (M), high (H) Completeness: PB=L, M = H, E = H Correctness: PB=H, M = H, E = H Currency: PB=L, M = L, E = M

adapted to meet user's needs while ensuring the efficient capture of complete, correct, and current data.

Author contributions

O.O.M. played a significant role in the creation and development of the research project. Along with R.I.B. and G.M.K., they collaborated in all stages of the study, including the design of methodology and refinement of data extraction. Y.Y. further improved the data analysis and extraction of evidence. O.O.M., R.I.B., and G.M.K. interpreted the results. F.D.S. and G.M.K. assisted with abstract and full-text review. O.O.M., G.M.K., and M.A. analyzed the search strategies and ensured reproducibility of the methodology. O.O.M. drafted the initial manuscript, with further revisions and completion by all authors. All authors were involved in critically analyzing the validity of the intellectual content in the final manuscript.

Supplementary material

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

Funding

This research was supported by the National Institute of Health (NIH), the National Institute of Nursing Research (NINR) grant award (R01 NR018416/NR/NINR NIH HHS/United States). The author is responsible for the content of this research and does not necessarily represent the official view of NIH or NINR.

Conflicts of interest

None declared.

Data availability

The data extracted from the 5 databases used in this review are available on reasonable request.

References

1. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med*. 2020;3(1):1-10. https://doi.org/10.1038/s41746-020-0221-y
2. Keenan GM, Lopez KD, Yao Y, et al. Toward meaningful care plan clinical decision support: feasibility and effects of a simulated pilot study. *Nurs Res*. 2017;66(5):388-398. https://doi.org/10.1097/NNR.0000000000000234
3. Fisher AM, Mtonga TM, Espino JU, et al. User-centered design and usability testing of RxMAGIC: a prescription management and general inventory control system for free clinic dispensaries. *BMC Health Serv Res*. 2018;18(1):703. https://doi.org/10.1186/s12913-018-3517-8

4. Huang AE, Hribar MR, Goldstein IH, Henriksen B, Lin WC, Chiang MF. Clinical documentation in electronic health record systems: analysis of similarity in progress notes from consecutive outpatient ophthalmology encounters. *AMIA Annu Symp Proc*. 2018;2018:1310-1318.
5. Keenan GM, Yakel E, Tschannen D, Mandeville M. Documentation and the nurse care planning process. In: Hughes RG, ed. *Patient Safety and Quality: An Evidence-Based Handbook for Nurses*. Advances in Patient Safety. Agency for Healthcare Research and Quality; 2008. Accessed February 16, 2022. http://www.ncbi.nlm.nih.gov/books/NBK2674/
6. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc*. 2013;20(1):144-151. https://doi.org/10.1136/amiajnl-2011-000681
7. Bian J, Lyu T, Loiacono A, et al. Assessing the practice of data quality evaluation in a national clinical data research network through a systematic scoping review in the era of real-world data. *J Am Med Inform Assoc*. 2020;27(12):1999-2010. https://doi.org/10.1093/jamia/ocaa245
8. Weiskopf NG, Bakken S, Hripcsak G, Weng C. A data quality assessment guideline for electronic health record data reuse. *EGEMS (Wash DC)*. 2017;5(1):14. https://doi.org/10.5334/egems.218
9. Kahn MG, Callahan TJ, Barnard J, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *EGEMS (Wash DC)*. 2016;4(1):1244. https://doi.org/10.13063/2327-9214.1244
10. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. *J Am Med Inform Assoc*. 1997;4(5):342-355.
11. Chen K, Hellerstein JM, Parikh TS. Designing adaptive feedback for improving data entry accuracy. In: *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*. UIST '10. Association for Computing Machinery; 2010:239-248. https://doi.org/10.1145/1866029.1866068
12. Bowman S. Impact of electronic health record systems on information integrity: quality and safety implications. *Perspect Health Inf Manag*. 2013;10(Fall):1c.
13. Cifuentes M, Davis M, Fernald D, Gunn R, Dickinson P, Cohen DJ. Electronic health record challenges, workarounds, and solutions observed in practices integrating behavioral health and primary care. *J Am Board Fam Med*. 2015;28(Suppl 1):S63-S72. https://doi.org/10.3122/jabfm.2015.S1.150133
14. Johnson SB, Bakken S, Dine D, et al. An electronic health record based on structured narrative. *J Am Med Inform Assoc*. 2008;15(1):54-64. https://doi.org/10.1197/jamia.M2131
15. Wilbanks BA, Berner ES, Alexander GL, Azuero A, Patrician PA, Moss JA. The effect of data-entry template design and anesthesia provider workload on documentation accuracy, documentation efficiency, and user-satisfaction. *Int J Med Inform*. 2018;118:29-35. https://doi.org/10.1016/j.ijmedinf.2018.07.006
16. Hasan S, Padman R. Analyzing the effect of data quality on the accuracy of clinical decision support systems: a computer simulation approach. *AMIA Annu Symp Proc*. 2006;2006:324-328.
17. The Pew Charitable Trusts. Poor Usability of Electronic Health Records Can Lead to Drug Errors, Jeopardizing Pediatric Patients. Published April 24, 2019. Accessed September 24, 2021. https://pew.org/2vb28et
18. 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. 290-09-00023I-7)*. Agency for Healthcare Research and Quality; 2011. Accessed September 24, 2021. https://digital.ahrq.gov/ahrq-funded-projects/electronic-health-record-usability-toolkit/citation/ehr-toolkit-background
19. Diaz-Garelli F, Strowd R, Lawson VL, et al. Workflow differences affect data accuracy in oncologic EHRs: a first step toward detangling the diagnosis data babel. *JCO Clin Cancer Inform*. 2020;4:529-538. https://doi.org/10.1200/CCI.19.00114
20. Nielsen J. Usability metrics: tracking interface improvements. *IEEE Softw*. 1996;13(6):1-2. https://doi.org/10.1109/MS.1996.8740869
21. Wronikowska MW, Malycha J, Morgan LJ, et al. Systematic review of applied usability metrics within usability evaluation methods for hospital electronic healthcare record systems. *J Eval Clin Pract*. 2021;27(6):1403-1416. https://doi.org/10.1111/jep.13582
22. Nielsen J. *Usability Metrics*. Nielsen Norman Group. Published January 20, 2001. Accessed December 1, 2021. https://www.nngroup.com/articles/usability-metrics/
23. Campbell EM, Sittig DF, Ash JS, Guappone KP, Dykstra RH. Types of unintended consequences related to computerized provider order entry. *J Am Med Inform Assoc*. 2006;13(5):547-556. https://doi.org/10.1197/jamia.M2042
24. Van Batavia JP, Weiss DA, Long CJ, et al. Using structured data entry systems in the electronic medical record to collect clinical data for quality and research: can we efficiently serve multiple needs for complex patients with spina bifida? *J Pediatr Rehabil Med*. 2018;11(4):303-309. https://doi.org/10.3233/PRM-170525
25. Walji MF, Kalenderian E, Tran D, et al. Detection and characterization of usability problems in structured data entry interfaces in dentistry. *Int J Med Inform*. 2013;82(2):128-138. https://doi.org/10.1016/j.ijmedinf.2012.05.018
26. Whittemore R, Knafl K. The integrative review: updated methodology. *J Adv Nurs*. 2005;52(5):546-553. https://doi.org/10.1111/j.1365-2648.2005.03621.x
27. Madandola OO, Keenan GM, Bjarnadottir RI, Dos Santos FC, Macieira TGR. *Impact of User Interface of Electronic Health Record on Data Quality: A Systematic Review*. PROSPERO; 2020. Accessed March 28, 2023. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020203998
28. Kmet LM, Lee RC, Cook LS; Alberta Heritage Foundation for Medical Research. Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields. Alberta Heritage Foundation for Medical Research; 2004. Accessed December 1, 2021. http://www.deslibris.ca/ID/200548
29. Adams WG, Conners WP, Mann AM, Palfrey S. Immunization entry at the point of service improves quality, saves time, and is well-accepted. *Pediatrics*. 2000;106(3):489-492. https://doi.org/10.1542/peds.106.3.489
30. Greenbaum NR, Jernite Y, Halpern Y, et al. Improving documentation of presenting problems in the emergency department using a domain-specific ontology and machine learning-driven user interfaces. *Int J Med Inform*. 2019;132:103981. https://doi.org/10.1016/j.ijmedinf.2019.103981
31. Urchek RJ, Morscher MA, Steiner RP, Adamczyk MJ. Orthopaedic resident use of an electronic medical record template does not improve documentation for pediatric supracondylar humerus fractures. *J Am Acad Orthop Surg*. 2019;27(8):e395-e400. https://doi.org/10.5435/jaaos-d-17-00818
32. Yang YS, Bass EJ, Bowles KH, Sockolow PS. Impact of home care admission nurses' goals on electronic health record documentation strategies at the point of care. *Comput Inform Nurs*. 2019;37(1):39-46. https://doi.org/10.1097/cin.0000000000000468
33. Kaka H, Ayearst R, Tran M, et al. Developing an IPAD® application for data collection in a Rheumatology Research Clinic. *Int J Technol Assess Health Care*. 2015;31(1-2):99-102. https://doi.org/10.1017/S0266462315000069
34. Skyttberg N, Chen R, Blomqvist H, Koch S. Exploring vital sign data quality in electronic health records with focus on emergency care warning scores. *Appl Clin Inform*. 2017;8(3):880-892. https://doi.org/10.4338/aci-2017-05-ra-0075
35. Jensen LG, Bossen C. Factors affecting physicians' use of a dedicated overview interface in an electronic health record: the importance of standard information and standard documentation. *Int J Med Inform*. 2016;87:44-53. https://doi.org/10.1016/j.ijmedinf.2015.12.009

36. Kostopoulou O, Tracey C, Delaney BC. Can decision support combat incompleteness and bias in routine primary care data? *J Am Med Inform Assoc*. 2021;28(7):1461-1467. https://doi.org/10.1093/jamia/ocab025
37. Kuru K, Girgin S, Arda K, Bozlar U. A novel report generation approach for medical applications: the SISDS methodology and its applications. *Int J Med Inform*. 2013;82(5):435-447. https://doi.org/10.1016/j.ijmedinf.2012.05.019
38. Avidan A, Weissman C. Record completeness and data concordance in an anesthesia information management system using context-sensitive mandatory data-entry fields. *Int J Med Inform*. 2012;81(3):173-181. https://doi.org/10.1016/j.ijmedinf.2011.12.009
39. Hua L, Gong Y. Toward user-centered patient safety event reporting system: a trial of text prediction in clinical data entry. *Stud Health Technol Inform*. 2015;216:188-192.
40. Brooke J. SUS: a "quick and dirty" usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, eds. *Usability Evaluation in Industry*. CRC Press; 1996. https://doi.org/10.1201/9781498710411
41. Skyttberg N, Chen R, Koch S. Man vs machine in emergency medicine – a study on the effects of manual and automatic vital sign documentation on data quality and perceived workload, using observational paired sample data and questionnaires. *BMC Emerg Med*. 2018;18(1):54. https://doi.org/10.1186/s12873-018-0205-2
42. Payne TH, Alonso WD, Markiel JA, Lybarger K, White AA. Using voice to create hospital progress notes: description of a mobile application and supporting system integrated with a commercial electronic health record. *J Biomed Inform*. 2018;77:91-96. https://doi.org/10.1016/j.jbi.2017.12.004
43. Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing. Federal Register. Published April 18, 2023. Accessed August 22, 2023. https://www.federalregister.gov/documents/2023/04/18/2023-07229/health-data-technology-and-interoperability-certification-program-updates-algorithm-transparency-and
44. Shakshuki EM, Reid M, Sheltami TR. An adaptive user interface in healthcare. *Procedia Comput Sci*. 2015;56:49-58. https://doi.org/10.1016/j.procs.2015.07.182
45. Lin YJ, Speedie S. Role-based and adaptive user interface designs in a teledermatology consult system: a way to secure and a way to enhance. *AMIA Annu Symp Proc*. 2003;2003:913.
46. Zheng K, Padman R, Johnson MP, Diamond HS. An interface-driven analysis of user interactions with an electronic health records system. *J Am Med Inform Assoc*. 2009;16(2):228-237. https://doi.org/10.1197/jamia.M2852
47. Gambino O, Rundo L, Cannella V, Vitabile S, Pirrone R. A framework for data-driven adaptive GUI generation based on DICOM. *J Biomed Inform*. 2018;88:37-52. https://doi.org/10.1016/j.jbi.2018.10.009
48. Hussain J, Ul Hassan A, Muhammad Bilal HS, et al. Model-based adaptive user interface based on context and user experience evaluation. *J Multimodal User Interfaces*. 2018;12(1):1-16. https://doi.org/10.1007/s12193-018-0258-2
49. Gajos KZ, Everitt K, Tan DS, Czerwinski M, Weld DS. Predictability and accuracy in adaptive user interfaces. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM; 2008:1271-1274. https://doi.org/10.1145/1357054.1357252