

# Application of human factors to improve usability of clinical decision support for diagnostic decision-making: a scenario-based simulation study

Pascale Carayon <sup>1</sup>, Peter Hoonakker,<sup>2</sup> Ann Schoofs Hundt,<sup>2</sup> Megan Salwei,<sup>1</sup> Douglas Wiegmann,<sup>1</sup> Roger L Brown,<sup>3</sup> Peter Kleinschmidt,<sup>4</sup> Clair Novak,<sup>5</sup> Michael Pulia,<sup>6</sup> Yudi Wang,<sup>4</sup> Emily Wirkus,<sup>7</sup> Brian Patterson<sup>6</sup>

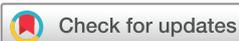
► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjqs-2019-009857>).

For numbered affiliations see end of article.

## Correspondence to

Dr Pascale Carayon, Department of Industrial and Systems Engineering, Wisconsin Institute for Healthcare Systems Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA; [pcarayon@wisc.edu](mailto:pcarayon@wisc.edu)

Received 28 May 2019  
Revised 11 October 2019  
Accepted 5 November 2019  
Published Online First  
27 November 2019



© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

**To cite:** Carayon P, Hoonakker P, Hundt AS, *et al.* *BMJ Qual Saf* 2020;**29**:329–340.

## ABSTRACT

**Objective** In this study, we used human factors (HF) methods and principles to design a clinical decision support (CDS) that provides cognitive support to the pulmonary embolism (PE) diagnostic decision-making process in the emergency department. We hypothesised that the application of HF methods and principles will produce a more usable CDS that improves PE diagnostic decision-making, in particular decision about appropriate clinical pathway.

**Materials and methods** We conducted a scenario-based simulation study to compare a HF-based CDS (the so-called CDS for PE diagnosis (PE-Dx CDS)) with a web-based CDS (MDCalc); 32 emergency physicians performed various tasks using both CDS. PE-Dx integrated HF design principles such as automating information acquisition and analysis, and minimising workload. We assessed all three dimensions of usability using both objective and subjective measures: effectiveness (eg, appropriate decision regarding the PE diagnostic pathway), efficiency (eg, time spent, perceived workload) and satisfaction (perceived usability of CDS).

**Results** Emergency physicians made more appropriate diagnostic decisions (94% with PE-Dx; 84% with web-based CDS;  $p<0.01$ ) and performed experimental tasks faster with the PE-Dx CDS (on average 96 s per scenario with PE-Dx; 117 s with web-based CDS;  $p<0.001$ ). They also reported lower workload ( $p<0.001$ ) and higher satisfaction ( $p<0.001$ ) with PE-Dx.

**Conclusions** This simulation study shows that HF methods and principles can improve usability of CDS and diagnostic decision-making. Aspects of the HF-based CDS that provided cognitive support to emergency physicians and improved diagnostic performance included automation of information acquisition (eg, auto-populating risk scoring algorithms), minimisation of workload and support of decision selection (eg, recommending a clinical pathway). These HF design principles can be applied to the design of other CDS technologies to improve diagnostic safety.

## BACKGROUND AND SIGNIFICANCE

Diagnostic errors have emerged as a critical patient safety issue.<sup>1–4</sup> Technological solutions such as clinical decision support (CDS) have the *potential* to improve diagnosis and reduce diagnostic errors.<sup>1 5</sup> However, in general, CDS adoption, use and impact have been limited,<sup>6</sup> partly because of poor usability and workflow integration.<sup>7–10</sup> Human factors (HF) methods and principles can enhance CDS usability, improve workflow integration and positively impact patient safety.<sup>8 11</sup> This is especially important in settings such as the emergency department (ED) where diagnostic decisions, for example, making a pulmonary embolism (PE) diagnosis, are often made under time pressure and in chaotic environments,<sup>12</sup> therefore leading to diagnostic errors.<sup>13</sup> In this study, we applied a HF-based approach to design a CDS to support PE diagnostic decision-making in the ED. The study objective was to assess whether a HF-based approach improves CDS usability, including cognitive support to PE diagnostic decision-making, efficiency, workload and satisfaction.

## CDS technologies for improving diagnosis

Health information technologies (IT), including CDS, can facilitate access to patient-related information and support diagnostic decision-making.<sup>1 5 14 15</sup> Several systematic reviews<sup>16–18</sup> indicate the *potential* for CDS technologies to improve ED diagnostic processes. However, the evidence on the benefits of CDS for ED

patient outcomes is rather weak,<sup>16 18 19</sup> often because the technologies are neither accepted nor used.<sup>20</sup> For instance, studies have demonstrated the positive (although small) impact of CDS on appropriate ordering of CT angiography (CTA) for PE diagnosis,<sup>18 21 22</sup> but many ED physicians (nearly 27% in one study<sup>21</sup>) may not use the CDS because of the time and data entry requirements. Achieving improvements in diagnostic safety with CDS implementation partly depends on the usability of CDS technologies and their integration into clinical workflow.<sup>15 23</sup> For PE diagnosis, CDS should support the diagnostic pathway and associated cognitive activities to avoid ordering wrong or unnecessary tests.<sup>22</sup> Indeed, suspected PE is a challenging diagnostic scenario in the ED, with the morbidity and mortality of delayed or missed diagnoses<sup>24 25</sup> needing to be weighed against potential harms of overtesting.<sup>26</sup> In a study of 328 patients receiving CTA for PE diagnosis,<sup>27</sup> emergency physicians indicated reasons for ordering a CTA: to confirm or rule out PE (93%), elevated D-dimers (66%) and fear of missing PE (55%). This latter reason (fear of missing PE) may represent a defensive behaviour, which can contribute to unnecessary radiation exposure.<sup>28</sup> Systematic risk assessment of patients using validated rules such as the Wells' score and the pulmonary embolism rule-out criteria (PERC) rule<sup>29</sup> is recommended so that PE is appropriately diagnosed in a timely manner<sup>30</sup> while avoiding unnecessary CTA scanning and radiation.<sup>28</sup> CDS can support the diagnostic process of emergency physicians when suspecting PE by suggesting appropriate clinical pathways, which are in concordance with these validated rules (eg, ordering a D-dimer test for patient with moderate PE risk as opposed to proceeding directly to CTA).

### HF-based design of health IT

Health IT, such as CDS, need to incorporate HF methods and principles to improve usability.<sup>31–34</sup> Clinical decision-making such as a diagnostic process involves cognitive activities of acquiring and analysing information and using this information to select a decision.<sup>1</sup> HF experts<sup>35</sup> recommend different levels of automation for cognitive activities at different information processing stages. High automation is useful for efficient information acquisition and analysis; the stage of decision selection should be supported by the CDS but not fully automated to preserve clinician judgement. Those HF design principles and usability heuristics, for example, consistency and visibility,<sup>36 37</sup> should be used in designing CDS. See [table 1](#) for the HF design principles applied to the CDS tested in this study.

Few studies have demonstrated the value of HF-based design in *actually improving the usability and safety* of CDS technologies. For instance, a simulation study<sup>38</sup> showed how the redesign of medication alerts based on HF design principles (eg, proximity compatibility)

improved usability, for example, enhanced efficiency and reduced prescription errors. Studies that incorporate HF methods and principles have focused on medication safety,<sup>38 39</sup> preventive care (eg, colorectal cancer screening<sup>40</sup>) and clinical reminders.<sup>41</sup> The value of HF-based design of CDS to support the diagnostic decision-making process has been suggested,<sup>1 42–44</sup> but has not been demonstrated yet.

### Study objectives

Using HF methods and principles, we created a CDS (known as PE-Dx CDS) to support the PE diagnostic pathway for ED physicians. The study objective was to assess whether the HF-based approach improved CDS usability and diagnostic decision-making as compared with a widely used web-based CDS. We assessed the new CDS (PE-Dx) and the web-based CDS on the three usability dimensions defined by the International Organisation for Standardisation<sup>45</sup> (effectiveness, efficiency and satisfaction) and tested the following hypotheses:

- ▶ Hypothesis 1 (effectiveness). Because PE-Dx provided cognitive support for the diagnostic clinical pathway, it would be more likely to lead to appropriate decision regarding diagnostic pathway, and greater confidence in decision as compared with the web-based CDS.
- ▶ Hypothesis 2 (efficiency). Because the PE-Dx was designed according to HF principles (eg, automation and workload minimisation), it would be more efficient and would lead to faster performance, fewer clicks and scrolls, and lower perception of workload as compared with the web-based CDS.
- ▶ Hypothesis 3 (satisfaction). Because the design of the PE-Dx CDS incorporated HF methods and principles, it would be more satisfying and produce a more positive score of perceived usability as compared with the web-based CDS.

## METHODS

### Organisational setting

We conducted an experimental study in collaboration with the ED of a large US academic teaching hospital, which has been using an EHR (Epic Systems) since 2008. The study was approved by the hospital's institutional review board. We used the STARE-HI guidelines to write this article.<sup>46 47</sup>

### Description of the CDS for PE diagnosis (PE-Dx CDS)

While there are several strategies for evaluating patients presenting with symptoms of PE, validated clinical guidelines both nationally and specifically in use in the study ED suggest a tiered approach using two risk scores: the Wells' score<sup>48</sup> and PERC rule.<sup>49 50</sup> These scores are available on websites and apps, for example, MDCalc, a widely used source for medical scores and algorithms ([www.mdcalc.com](http://www.mdcalc.com)). Patients are initially screened with the Wells' score. If the Wells' score is higher than 6, PE risk is high and cross-sectional imaging

**Table 1** HF design of PE-Dx CDS

HF design process	Design activities and outcomes
1. HF analysis of PE diagnosis in the ED	Identification of multiple work system barriers and facilitators to cognitive work involved in PE diagnosis; in particular technology-related barriers for information gathering activities in PE diagnostic process
2. Analysis of suggestions made by ED physicians regarding health IT solutions	Main suggestions for CDS to support PE diagnosis in ED included: <ul style="list-style-type: none"> <li>▶ Integration with EHR</li> <li>▶ Auto-population</li> <li>▶ Automated scoring</li> <li>▶ Support for decision regarding order.</li> </ul>
3. Nine interdisciplinary design sessions over 1 year for a total of 13.5 hours (each session was between 1 and 2 hours)	Decision by design team to focus the CDS on how to support diagnostic pathway for PE diagnosis. Multiple disciplines involved in design sessions: HF, emergency medicine and software engineering/programming. Use of iterative design process with multiple mock-ups of PE-Dx CDS (created in PowerPoint)
4. Two 1-hour focus groups with four ED physicians (interspersed between design sessions) for review of PE-Dx CDS mock-up	Confirmation of preference for two-step CDS: presentation of Wells' and then PERC; match with hospital policy. Use of highest heart rate/pulse value and lowest O <sub>2</sub> saturation value; but need to allow physicians to update these values
5. Creation of PE-Dx CDS in EHR Playground	Programming of PE-Dx CDS in EHR Playground based on PowerPoint mock-up
6. Two-hour participatory heuristic evaluation of PE-Dx CDS with seven participants (HF engineers, physicians, systems analyst)	Identification of 19 usability problems; 13 problems resolved through design changes to PE-Dx CDS; six problems not addressed because of constraints of flowsheet functionality of local EHR Production of final PE-Dx CDS tested in usability evaluation

ED, emergency department; IT, information technology; PE-Dx CDS, clinical decision support for pulmonary embolism diagnosis.

HF design principles	Implementation in PE-Dx CDS
Automation of information acquisition <sup>35</sup>	Auto-population of some Wells' and PERC criteria using EHR data
Automation of information analysis <sup>35</sup>	Computation of Wells' score by CDS
Support of decision selection <sup>35</sup>	Provision of recommendation for next step in diagnostic pathway, for example, doing nothing, ordering D-dimer, ordering CTA scan
Explicit control/flexibility <sup>35 36</sup>	Ability to change values for Wells' criteria—for example, possible to change heart rate/pulse—to support physician clinical judgement and unique patient situations
Minimisation of workload <sup>36</sup>	Minimisation of data entry; for example, data for Wells' automatically populated in PERC PERC appears only if Wells' score is low No need to enter data for all PERC criteria once any PERC criterion is positive Automatic generation of text for documentation of medical decision-making
Consistency <sup>36 37</sup>	Consistency with how information is presented in other parts of the EHR; for example, use of Yes/No toggle Consistency with how Wells' and PERC criteria are listed on MDCalc website that is routinely used by physicians
Chunking/grouping <sup>36 74</sup>	Wells' criteria and PERC rule presented separately Placement of CDS in ED navigator of EHR
Visibility <sup>37</sup>	Indication of points/weights assigned to each Wells' criterion to make it clear/transparent how Wells' score is computed
Error prevention <sup>36 37</sup>	To avoid documenting wrong Wells' score, all Wells' criteria must be addressed

HF, human factors; PERC, pulmonary embolism rule-out criteria.

such as CTA should be ordered. If the Wells' score is between 2 and 6, PE risk is moderate and guidelines recommend ordering a D-dimer (a blood test which, when normal, effectively excludes PE in intermediate to low-risk scenarios, and if positive prompts cross-sectional imaging). If the Wells' score is lower than 2, PE risk is low. In this case, it is appropriate to apply the PERC. If any PERC criteria are positive, a D-dimer is recommended; if all PERC criteria are negative, a PE diagnosis can be excluded without any further work-up.<sup>29 49 51</sup> An objective of this staged diagnostic process recommended by the American College of Physicians<sup>49</sup> is to reduce patients' exposure to radiation.

Using a HF-based approach, we created PE-Dx, a computerised CDS integrated in the EHR to support the PE diagnostic pathway. The human-centred design process relied on multiple, iterative steps that combined work system analysis, design sessions, focus groups and a heuristic evaluation,<sup>52</sup> and incorporated HF design principles (table 1).

Using the flowsheet functionality in Epic, a clinical systems analyst of the participating hospital programmed the PE-Dx CDS based on a PowerPoint mock-up developed in the HF design process (table 1). PE-Dx incorporated Wells' criteria for PE and the PERC rule in accord with hospital policy and

established guidelines.<sup>49</sup> The clinician first considered the results of the Wells' criteria and, when a patient was 'Wells' low', then s/he completed PERC. Data available in the EHR were used to auto-populate the PE-Dx CDS. Some auto-populated fields could be overridden by physicians (ie, heart rate and O<sub>2</sub> saturation). A recommendation regarding the diagnostic pathway appeared when the final criterion was completed. If the Wells' score was 'low' (ie, <2), the auto-populated PERC criteria appeared. An affirmative response to any of the PERC criteria concluded the process and a recommendation appeared to order a D-dimer. See figure 1 for screenshots of the CDS, and table 2 for a comparison of the usual diagnostic process (MDCalc) and the PE-Dx CDS process.

### Study design

We conducted a scenario-based simulation study with a repeated measures design. Participants were first exposed to the web-based CDS used in clinical practice (MDCalc; www.mdcalc.com) and then to the PE-Dx CDS in two separate sessions, each lasting approximately 30 min and occurring about 1 week apart. Five randomly ordered patient scenarios were presented during each session (online supplementary appendix for scenario details). The scenarios represented the range of decisions (no further action, order D-dimer, order CTA scan) that physicians make when suspecting PE, and all possible combinations of the Wells' score and the PERC rule.<sup>49</sup> A physician on the study team drafted the scenarios, which were pilot tested with two other physicians. The scenarios were designed such that objectively presented data placed patients in one of five possible pathways, each with an appropriate decision:

1. Wells' low (<2) / PERC negative; appropriate decision=no further action required
2. Wells' low (<2) / PERC positive; appropriate decision=order D-dimer
3. Wells' low (<2) / PERC positive due solely to age >49; appropriate decision=order D-dimer
4. Wells moderate ( $\geq 2$  and  $\leq 6$ ), PERC inappropriate; appropriate decision=order D-dimer
5. Wells' high (>6), PERC inappropriate; appropriate decision=order CTA scan.

Differences existed between the scenarios of session 1 and those of session 2 to minimise recall; differences were minor and not relevant for PE diagnosis. The creation of the scenarios integrated best practices outlined by Russ and Saleem.<sup>53</sup>

### Participants

In all, 24 emergency medicine (EM) residents (eight from each of three residency classes) and eight EM faculty participated in the study. The study took place over a 10-week period from April through June 2018. Results of a power analysis showed that, for a power of 0.80 with five different scenarios, we needed a sample

size of at least 21 participants to obtain an effect size of 0.70 with an alpha of 0.05; this was equivalent to recruiting six physicians in each of the four targeted physician groups (three residency years and attendings). To ensure sufficient power and given the uncertainty around completion of both sessions by participants, we recruited eight participants in each group, for a total of 32 physicians. Physicians were recruited via email, and all presented for both sessions. Total compensation of \$100 was provided: \$20 after session one and \$80 after session 2.

### Study procedures

For study purposes, the PE-Dx CDS ran in the Epic 'Playground', an electronic environment that mimics the actual electronic health record (EHR) with fictitious patients and is used for testing and training. Following a precise script (available on request), trained researchers explained to the study participants the goal of the experiment, that is, 'to evaluate a new clinical decision support tool—CDS—to facilitate making decisions about PE'. They then directed the study participants to complete the following tasks for each patient scenario: (1) review prepopulated ED notes containing relevant clinical information on the patient, (2) complete MDCalc (Wells' and PERC when necessary) in session 1 and the PE-Dx CDS in session 2, (3) indicate on the paper survey what clinical pathway they would follow (order nothing, D-dimer or CTA) and their level of confidence with the decision, and (4) complete an electronic survey. After completing the final scenario, the electronic survey included additional questions. On completion of session 2, we conducted a short debrief interview to ask participants' preference for either MDCalc or PE-Dx and perceived barriers and facilitators for implementing PE-Dx. Comments by participants during debriefing are used as quotes when discussing study results.

### Data collection and measurement

We used multiple data collection methods to assess usability: screen capture software to video record task performance (Camtasia), stopwatch, paper survey and electronic survey.

#### Screen capture software

The screen capture software (Camtasia) started at the beginning of the session and recorded continuously every activity performed by participants until the end of the session. Researchers reviewed the video recordings and extracted data to assess:

- ▶ Number of clicks for each scenario. Counting of clicks used to navigate or choose/select an item began once the participant typed the patient's name and stopped when the participant documented his/her decision.
- ▶ Number of scrolls for each scenario. Counting of uninterrupted scrolls on a page or side bar started and ended similar to the counting of clicks.

A.

Hyperspace - UWH ED - PLY-Playground

Antwerp, Deacy  
Female, 62 year old, 05/15/1956

Chief Complaint: Dyspnea

Wells' Criteria

Clinical signs and symptoms of DVT	3+Yes	0+No
PE is #1 diagnosis OR equally likely	3+Yes	0+No
HR/Pulse > 99 (1.5+Yes, 0+No)		
Immobilization at least 3 days OR surgery in the previous 4 weeks	1.5+Yes	0+No
Previous, objectively diagnosed PE or DVT	1.5+Yes	0+No
Hemoptysis	1+Yes	0+No
Malignancy with treatment within 6 months or palliative	1+Yes	0+No

Wells' Criteria Score: 0.0

Wells' Criteria Recommendation: None

Documentation for ED Provider Note

Click to generate recommended ED Provider Note documentation.

The documentation above will automatically populate in your ED Provider Note. If your note has already been started, refresh the note contents to bring in this documentation. If a D-dimer or imaging order is recommended and has not yet been ordered, after clicking Next or Close you will be presented with a BPA to place the appropriate order.

© 2019 Epic Systems Corporation. Used with permission.

B.

Hyperspace - UWH ED - PLY-Playground

Antwerp, Dean  
Male, 46 year old, 05/15/1972

Chief Complaint: Chest Pain, Cough and/...

PERC Criteria

Prior PE or DVT (autopopulated from Wells' Criteria)	No
HR/Pulse > 99 (autopopulated from Wells' Criteria)	No
Hemoptysis (autopopulated from Wells' Criteria)	No
Age > 49	46
SaO2 on room air < 95%	99
Unilateral leg swelling	Yes/No
Recent surgery or trauma (requiring intubation or hospitalization in previous 4 weeks)	Yes/No
Hormone use	Yes/No

PERC Result: Negative

PERC Recommendation: None

Documentation for ED Provider Note

Click to generate recommended ED Provider Note documentation.

Patient is at low risk for pulmonary embolism based on Wells' Score <2 and is PERC negative. Given these findings, no further screening for pulmonary embolism was performed.

The documentation above will automatically populate in your ED Provider Note. If your note has already been started, refresh the note contents to bring in this documentation. If a D-dimer or imaging order is recommended and has not yet been ordered, after clicking Next or Close you will be presented with a BPA to place the appropriate order.

© 2019 Epic Systems Corporation. Used with permission.

**Figure 1** PE-Dx CDS: wells' criteria, wells' score and recommendation (A); PERC criteria (B). PE-Dx CDS, clinical decision support for pulmonary embolism diagnosis; PERC, pulmonary embolism rule-out criteria.

- ▶ Number of navigation elements used for each scenario. Navigation elements included windows in the EHR (eg, ED provider note, medication list, medical history) and browser windows (eg, MDCalc for Wells') that physicians interacted with during navigation.

#### Stopwatch

Time to complete each scenario was measured with a stopwatch on a smart phone. Start time was defined as the moment that the participant started typing the patient's name in ED dashboard. End time was defined

**Table 2** Comparison of diagnostic decision pathway with MDCalc and PE-Dx CDS

	MDCalc (web-based CDS)	PE-Dx CDS
<b>Integration with EHR</b>		
	Not integrated in EHR; separate website	Integrated in EHR
<b>Wells' score (seven criteria)</b>		
Initiation	User needs to open browser, type MDCalc PE in search engine in browser and click on URL for MDCalc website	User clicks on PE-Dx button in ED Navigator
Criteria responses	User does not have to answer all (7) criteria before obtaining a score	User has to address all (7) criteria before obtaining the Wells' score
Auto-population	No criteria are auto-populated; therefore, the user may need to navigate between the EHR and MDCalc to search for relevant information	One criterion (heart rate >100) is auto-populated; user may need to navigate within the EHR to search for other relevant information
<b>PERC rule (eight criteria)</b>		
Initiation	User must access a different website within MDCalc to review PERC criteria	PERC is integrated in CDS and becomes active when indicated per hospital policy
Criteria responses	User must respond yes to only one criterion to receive PERC-positive result	User must respond yes to only one criterion to receive PERC-positive result
Auto-population	Does not auto-populate, nor copy results from Wells' criteria. Therefore, user may need to navigate between the EHR and MDCalc to search for relevant information	Auto-populates three questions (age, heart rate, O <sub>2</sub> saturation) and copies forward responses from comparable Wells' criteria (prior PE or deep vein thrombosis, hemoptysis). Therefore, user must address only three criteria (recent trauma or surgery, exogenous oestrogen, unilateral leg swelling)
<b>Risk assessment and decision-making</b>		
Result/score	Provides the user with a score (eg, 3.0 points for Wells' score) and compares that score with the population (eg, moderate-risk group: 16.2% chance of PE in ED population)	Provides the user with a score
Recommendations/next steps	Provides the user with suggestion for next steps; suggestion uses two different models, which can be confusing	Provides the user with a single recommendation in accord with hospital policy and established guidelines <sup>49</sup>
Order *	Does not automatically provide option to enter an order	Provides the option to go to CPOE to order D-dimer or CTA scan when applicable
Documentation *	Allows user to copy results (summary and scores on individual questions) into an EHR note, but user must register to use the feature	Allows the user to populate the results and recommendation directly into his/her note

\*The automated creation of an order and documentation of the decision were not tested in this study, but are available in the implemented version of PE-Dx CDS.

CPOE, computerized provider order entry; CTA, CT angiography; ED, emergency department; PE, pulmonary embolism; PE-Dx CDS, clinical decision support for pulmonary embolism diagnosis; PERC, pulmonary embolism rule-out criteria.

as the moment that the participant made his/her choice for the diagnostic pathway and recorded it on the paper survey.

**Paper survey after each scenario**

After completing each scenario, participants filled out a two-question paper survey. The first question asked for their decision regarding the diagnostic pathway: order nothing, a D-dimer test or a CTA. Their decision was compared with the guideline-supported decision for that particular scenario; this produced the measure of decision appropriateness. In the second question, participants indicated their confidence in their decision about the diagnostic pathway, measured on a 100-mm visual analogue scale, ranging from 0 (no confidence at all) to 100 (very high confidence).

**Electronic survey after each scenario**

After completing the paper survey, participants filled out a short questionnaire on an iPad, programmed in

the Qualtrics web-based survey software. The NASA Task Load (TLX) questionnaire<sup>54</sup> measured six dimensions of workload on a scale from 1 (low) to 10 (high): mental workload, physical workload, temporal workload, effort, frustration and performance. We adapted the question on physical workload to address physical activities of computer work, replacing the original text about pushing, pulling, turning, controlling, activating, etc with text on navigating, scrolling, clicking, etc.

**Electronic survey at the end of each session**

At the end of each session, participants answered 17 questions from the Computer System Usability Questionnaire (CSUQ).<sup>55</sup> The original CSUQ has 19 questions, but two questions about error messages and online help were removed because MDCalc and PE-Dx did not have those features. The Cronbach's alpha for the 17-item CSUQ was 0.95 in our study, which is similar to the reliability score in the original CSUQ study.<sup>55</sup> The electronic survey used at the end

of each scenario and at the end of each session can be found at [https://cqpi.wiscweb.wisc.edu/wp-content/uploads/sites/599/2018/12/PE\\_DX\\_Questionnaire\\_Session\\_2\\_Final1.pdf](https://cqpi.wiscweb.wisc.edu/wp-content/uploads/sites/599/2018/12/PE_DX_Questionnaire_Session_2_Final1.pdf).

### Data analysis

All outcome measures were transformed for normalcy in the models used for data analysis, and for the marginal graphics were back transformed to the original metrics for easier interpretation. Data analysis was conducted using a three-level (physician-session-scenario) empirical Bayesian (EB) regression model, allowing us to model the pre-post and scenario dependencies. We used a Bayesian estimation framework to analyse our models. Bayesian methods are a useful alternative to iterative generalised least squares estimation in multilevel regression models, particularly with smaller samples.<sup>56</sup> To obtain the benefits of Bayesian methods, a prior distribution must be informative.<sup>57</sup> We estimated our effects with EB Markov Chain Monte Carlo estimates,<sup>58 59</sup> using data-derived estimates from our dataset as our EB informative priors.<sup>56 58 60</sup> The Bayesian estimates and the analysis of our multilevel models were constructed using MLwiN V.3.02. We used the three-level EB model to obtain the estimates of marginal means and report the comparison of outcome measures for session 1 (MDCalc) and session 2 (PE-Dx).

## RESULTS

### Participant characteristics

Eight of the 32 (25%) study participants were women and most (88%) were between 25 and 34 years old. Study participants used MDCalc primarily on the

computer (75%) or their phone (25%). Most study participants used Wells' score (81%) or PERC rule (91%) on MDCalc. All participants completed both sessions of the study.

### Effectiveness

The percentage of appropriate decisions increased from 84% using MDCalc to 94% with the PE-Dx CDS ( $p < 0.01$ ) (table 3). Physician confidence in their decision about the diagnostic pathway did not change with PE-Dx; confidence was high for both CDS (mean around 80 on a 0–100 scale).

### Efficiency

Efficiency was measured with five variables; four of them were favourably impacted by the HF-based CDS and one variable was worse with PE-Dx (table 3). Physicians spent significantly less time per scenario with PE-Dx (117s with MDCalc vs 96s with PE-Dx;  $p < 0.001$ ). The difference of means was 20s, which represented about 20% of time saved per scenario. Physicians used fewer scrolls per scenario (seven with MDCalc vs six with PE-Dx;  $p < 0.001$ ) and interacted with fewer interface elements with the HF-based CDS (16 elements with MDCalc vs 10 elements with PE-Dx;  $p < 0.001$ ). There were significant differences in the NASA TLX global and individual scores (except for performance) between the two CDS (figure 2). PE-Dx was associated with lower overall perceived workload, as well as lower scores on mental workload, physical workload, temporal workload, effort and frustration. The number of clicks per scenario increased slightly with PE-Dx (16 with MDCalc vs 18 with PE-Dx;  $p < 0.01$ ).

Dependent variables	MDCalc	PE-Dx CDS	Effect size (CI)	P value
<b>Effectiveness</b>				
% appropriate decision	83.75% (134/160)	94.38% (151/160)	0.35* (0.32 to 0.37)	$p < 0.01$
Confidence level†	80.21 (18.94)	82.71 (18.21)	0.13‡ (−0.05 to 0.31)	$p = 0.14$
<b>Efficiency</b>				
Time per scenario (in seconds)	117.37 (38.91)	95.84 (25.42)	−0.55‡ (−0.71 to −0.38)	$p < 0.001$
Number of clicks per scenario	16.49 (4.82)	17.90 (3.91)	0.29‡ (0.09 to 0.48)	$p < 0.01$
Number of scrolls per scenario	7.47 (2.92)	6.31 (2.50)	−0.39‡ (−0.57 to −0.21)	$p < 0.001$
Number of navigation elements per scenario	15.53 (5.01)	10.23 (2.74)	−1.05‡ (−1.26 to −0.83)	$p < 0.001$
Perceived workload¶	4.45 (1.53)	3.64 (1.78)	−0.52‡ (−0.68 to −0.37)	$p < 0.001$
<b>Satisfaction§</b>				
CUSQ: overall satisfaction	5.62 (0.94)	6.18 (0.66)	0.58‡ (0.18 to 0.98)	$p < 0.001$
CUSQ: system usefulness	5.65 (0.95)	6.26 (0.67)	0.62‡ (0.19 to 1.05)	$p < 0.01$
CUSQ: information quality	5.73 (0.98)	6.17 (0.73)	0.43‡ (0.03 to 0.83)	$p < 0.05$
CUSQ: interface quality	5.48 (1.15)	6.02 (0.86)	0.45‡ (0.07 to 0.84)	$p < 0.05$

\*Cohen (h) effect size based on the inverse sine of the square root of the proportional values.

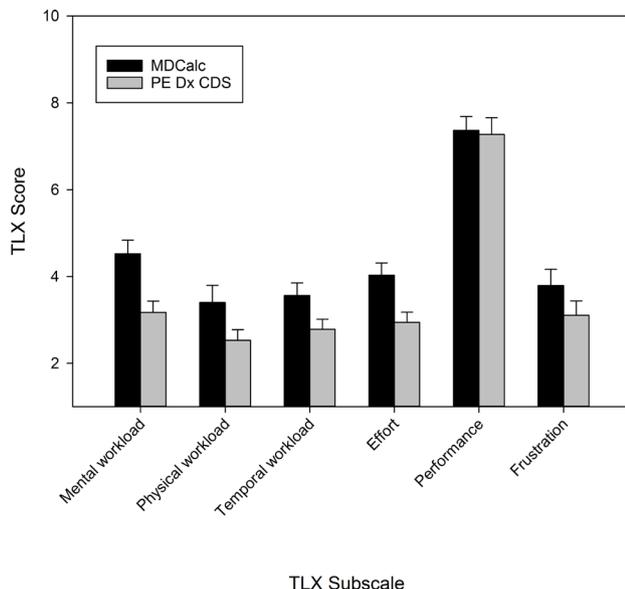
†One question: 0 (no confidence) to 100 (very high confidence).

‡Becker<sup>75</sup> adjusted effect size for repeated measures.

§Response categories: 1 (strongly disagree) to 7 (strongly agree).<sup>55</sup>

¶Scale: 1 (low) to 10 (high).<sup>54</sup>

CUSQ, computer system usability questionnaire; PE-Dx CDS, clinical decision support for pulmonary embolism diagnosis.



**Figure 2** Impact of PE-Dx CDS on perceived workload. Results are shown as mean±SEM. Higher scores indicate higher perceived workload. The differences on the NASA-TLX subscales were statistically significant, except for performance. PE-Dx CDS, clinical decision support for pulmonary embolism diagnosis.

### Satisfaction

The CUSQ produced one overall satisfaction score, and three separate scores of (1) system usefulness, (2) information quality and (3) interface quality. The HF-based CDS performed better on all four measures (table 3). Physicians were more satisfied with the usability of PE-Dx compared with MDCalc, including overall perceived usability, usefulness (eg, ‘I can effectively complete my work using this system’), information quality (eg, ‘It is easy to find the information I needed’) and interface quality (eg, ‘The interface of this system is pleasant’).

In all, 28 of the 32 participating physicians (88%) indicated their preference for PE-Dx. One physician would continue using MDCalc as s/he may not agree with the recommendation of PE-Dx. Three other physicians were unsure which CDS they preferred.

### DISCUSSION

In this study, we used HF methods and principles to design a CDS that supports the PE diagnostic process and compared this CDS (PE-Dx) to an existing web-based CDS (MDCalc). Results demonstrated that the HF-based CDS produced better usability and improved the PE diagnostic process. The majority of our hypotheses were supported: 9 of the 11 variables of usability improved with PE-Dx (table 3). Improvements in both objective (eg, time) and subjective (eg, satisfaction) measures of usability were achieved with PE-Dx. Physicians reported greater satisfaction with PE-Dx, including usefulness, information quality and interface quality (table 3). After being exposed to it,

a majority of physicians (88%) indicated their preference for PE-Dx.

### Cognitive support of PE-Dx CDS to PE diagnostic decision-making

PE-Dx was more effective than MDCalc as it provided cognitive support for the PE diagnostic pathway: 94% of physicians chose the appropriate clinical pathway when using PE-Dx (vs 84% with MDCalc). With MDCalc, physicians could decide not to follow the sequence of computing the Wells’ score first and then looking at PERC if appropriate. In a follow-up analysis,<sup>61</sup> we found that study participants followed the recommended workflow almost always with PE-Dx (98%), and only half of the time with MDCalc (51%). PE-Dx was built to guide physicians in a clinical workflow that started with the Wells’ score and continued with the PERC rule; this led to a clear recommendation to either do nothing, order a D-dimer or order a CTA.

PE-Dx helped physicians make a decision about which clinical pathway to take (HF design principle of support for decision selection).<sup>35 43</sup> However, in nine cases (out of 160 cases or 6%), physicians did not make the appropriate decision. In one case, the study participant made a data entry error, which led to a recommendation different from the one intended for this specific scenario. In seven cases, study participants chose no further testing, whereas PE-Dx recommended a D-dimer as the Wells’ score was below two with PERC positive; in one case, the physician ordered a CTA, whereas PE-Dx suggested a D-dimer. It is possible that in those eight cases (out of 160 cases) physicians disagree with the recommendation from the CDS or did not trust the CDS.<sup>62</sup> But we are unsure why study participants made those decisions as we did not ask them to think aloud so that we could obtain accurate estimates of time needed to complete tasks.

Using clinical decision and prediction rules has been identified as a health IT functionality to support making a diagnosis, ordering tests and determining a diagnostic plan<sup>5</sup>; this is what PE-Dx does. PE-Dx does not fully automate the decision selection, but provides a recommendation to the physician regarding next steps, that is, doing nothing, ordering a D-dimer or ordering a CTA. This partial automation of decision selection is a HF design principle.<sup>35</sup> High automation at the stages of information acquisition and analysis is another HF design principle applied to PE-Dx: some criteria were auto-populated (information acquisition) and scores were automatically calculated (information analysis) (table 1). Physicians liked this PE-Dx CDS functionality as indicated in a debrief interview: *I like that it pulls the vital signs in particular and also is smart enough to look at the entire vital signs from the entire encounter, another area in which I do feel more confident; that a computer would be able to do that more reliably than me with less error long-term.*

### Saving time and reducing workload with PE-Dx CDS

Physicians completed the scenarios about 20% faster with PE-Dx, and used fewer scrolls and navigation elements (eg, browser windows, EHR elements). During a debrief interview, a resident confirmed the importance of saving time with PE-Dx: *It took the criteria from the Wells' score that matched the PERC score and just auto-populated that as well, which is extremely, you know, efficient. That would help the workflow a lot. Instead of having to go back and forth in MDCalc and trying to type in the numbers between MDCalc and Epic, it would be much quicker.* However, we found that PE-Dx increased the number of clicks per scenario. On the one hand, when using PE-Dx a physician needs fewer clicks because s/he does not have to leave the EHR, open a browser, and go back and forth between the EHR and MDCalc website (table 2). On the other hand, PE-Dx requires more clicks as the physician needs to provide information for all seven Wells' criteria, whereas MDCalc does not require the user to answer all questions before calculating the Wells' score (table 2). We made the decision to ask physicians to address all Wells' criteria to avoid wrong decisions (principle of 'error prevention' in table 1). Overall, the PE-Dx CDS had a slightly higher number of clicks per scenario than MDCalc (table 3). Even though PE-Dx slightly increased clicks, physicians reported lower workload, except for the performance dimension (figure 2).

Several HF design principles contributed to the impact of PE-Dx on efficiency, in particular auto-population and minimising workload. For the Wells' score, only one criterion was auto-populated from the EHR, and for PERC, six criteria were auto-populated. Future improvement in the PE-Dx CDS should include additional auto-population, for example, using natural language processing. A benefit of PE-Dx was to use data entered in the Wells' score to populate the PERC rule; this reduced data entry. As indicated above, automation of information acquisition is a HF design principle,<sup>35</sup> which was applied to PE-Dx. This was confirmed as a good approach by a participating physician: *It is great that it is all within (the EHR)... and easily accessible with one click. That it auto-populates is a huge function. To be able to take out that step of having to manually look up information and input information,... I felt (it) sped things up entirely and also made me more confident in the sense that it took out my own doubts about my ability to consistently transcribe information.*

The performance subscale of the NASA TLX produced relatively high scores (means around 7 on a 1–10 scale) for both CDS. Physicians perceived high satisfaction with their performance on the experimental tasks; this matched the data on confidence in decision that was also high and similar for both CDS (table 3).

### Evidence for value of HF-based design

Our study provides empirical evidence for the value of HF-based design in increasing usability of health IT, such as CDS, and improving diagnostic decision-making. We showed that a human-centred design process integrating multiple HF methods and design principles (table 1) produced a more usable CDS that provided cognitive support for the diagnostic process. The HF design principles implemented in PE-Dx (table 1), in combination with other HF criteria such as those in heuristic usability evaluations,<sup>36 37</sup> could be applied to other CDS, in particular CDS targeting diagnostic processes that involve cognitive activities of information gathering, information integration and interpretation, and working diagnosis.<sup>1</sup>

The HF design principles can address the information fragmentation and overload of current EHR technologies, which contribute to physicians' frustration, stress and burnout.<sup>63</sup> Our HF-based CDS improved efficiency (eg, faster performance, lower perceived workload), although the number of clicks per scenario was slightly higher with PE-Dx (table 3). These positive results were likely related to auto-population, minimisation of data entry and chunking/grouping, that is, HF design principles aligned with suggestions made by ED physicians early in the design process (table 1). Therefore, a HF-based design that involves analysing the actual work of clinicians, participation from clinicians in the design process and implementation of HF design principles can enhance the usability of health IT such as CDS. As data continue to emerge about poor usability and workflow integration of health IT and adverse patient safety consequences,<sup>64</sup> it is essential to integrate HF methods and principles in technology design processes.<sup>11 65–68</sup> This is particularly important with the emergence of artificial intelligence, which can be used to design CDS technologies that complement and enhance clinical decision-making.<sup>69–71</sup>

### Strengths and weaknesses of the study

Findings of this scenario-based simulation study may not extend to the actual implementation of PE-Dx. When a technology is implemented in the real-work environment, challenges may occur that affect technology acceptance and use. Oftentimes, challenges are related to problems with technology usability and workflow integration, and misfit of the technology with the rest of the work system.<sup>72</sup> As our CDS technology was designed according to HF methods and principles, its usability was rated positively in our experimental study. But implementation issues may still arise. Therefore, future research should evaluate the actual implementation of PE-Dx, its use in the clinical environment, and its impact on key process and outcome measures, for example, appropriate use of lab or imaging tests. This research should use qualitative and quantitative research methods,<sup>73</sup> as well as objective and subjective measures of usability.<sup>15</sup> This

follow-up research can assess two functionalities of PE-Dx that were not tested in the simulation study: automated creation of lab test order (D-dimer) or imaging study (CTA) in computerized provider order entry (CPOE), and automated documentation of decision in ED provider note. Because we focused on cognitive support provided in the diagnostic pathway, the experimental setup and data collection stopped right after the physician made a diagnostic decision regarding next steps (order nothing, D-dimer or CTA).

Because the study took place in a single ED of a US academic hospital, it is difficult to generalise the results to other settings and emergency physicians. A strength of the study was its experimental approach; therefore, differences between the sessions are likely attributable to differences in the two CDS technologies (MDCalc and PE-Dx). Whereas scenarios were presented in a random order during each session, the order of CDS presentation was not counterbalanced: participants were first exposed to their current CDS (MDCalc) in session 1 and to PE-Dx in session 2 to assess the incremental impact of using the new HF-based CDS compared with usual practice. The lack of counterbalancing of conditions could affect our results if the participants' performance improved over time with their involvement in the study, which is unlikely. Further bias may have occurred because neither participants nor researchers were blinded; however, this is unlikely as we used a repeated measures design and implemented study procedures with detailed scripting of instructions given to participants by trained researchers. Another strength is that improvements were shown on all usability dimensions: with the newly designed CDS, the diagnostic pathway of PE was more effective and efficient, and physicians perceived lower workload and were more satisfied with the technology.

## CONCLUSION

Ensuring that HF principles are incorporated into CDS design is critical to improve CDS usability, which includes supporting appropriate diagnostic pathway, saving time, reducing perceived workload and improving physician satisfaction with the technology. This research suggests that HF principles such as automating information acquisition (eg, auto-population) and information analysis (eg, computing risk scores), minimising workload (eg, minimising data entry) and supporting decision selection are key to improving CDS usability and consequently diagnostic safety. Clinicians report greater satisfaction with a CDS designed according to HF principles. Future research should examine the effectiveness of these HF principles for other health IT and assess the implementation of PE-Dx and its clinical impact.

## Author affiliations

<sup>1</sup>Department of Industrial and Systems Engineering, Wisconsin Institute for Healthcare Systems Engineering, University of Wisconsin-Madison, Madison, Wisconsin, USA

<sup>2</sup>Center for Quality and Productivity Improvement, University of Wisconsin-Madison, Madison, Wisconsin, USA

<sup>3</sup>School of Nursing, University of Wisconsin-Madison, Madison, Wisconsin, USA

<sup>4</sup>Department of Medicine, University of Wisconsin-Madison, Madison, Wisconsin, USA

<sup>5</sup>UW Health, Madison, Wisconsin, USA

<sup>6</sup>Department of Emergency Medicine, University of Wisconsin-Madison, Madison, Wisconsin, USA

<sup>7</sup>Department of Population Health Sciences, University of Wisconsin-Madison, Madison, Wisconsin, USA

**Correction notice** The article has been corrected since it was published online first. Some minor changes have been done in table 3.

**Acknowledgements** We would like to thank the residents and attending physicians who participated in the study.

**Contributors** PC, ASH, PH, BP and DW designed the study. All authors were involved in the human-centered design process, data collection and data analysis. All authors reviewed the manuscript before submission.

**Funding** This research was made possible by funding from the Agency for Healthcare Research and Quality (AHRQ), Grant Numbers: R01HS022086-Principal Investigator: Pascale Carayon, and K08HS024558-Principal Investigator: Brian Patterson; and was supported by the Clinical and Translational Science Award (CTSA) program, through the NIH National Center for Advancing Translational Sciences (NCATS), Grant Number: 1UL1TR002373. The content is solely the responsibility of the authors and does not necessarily represent the official views of the AHRQ or NIH.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** No data are available.

## ORCID iD

Pascale Carayon <http://orcid.org/0000-0003-4632-6930>

## REFERENCES

- Balogh EP, Miller BT, Ball JR. *Improving diagnosis in health care*. Washington, DC: National Academies Press, 2015.
- Bates DW, Singh H. Two Decades Since *To Err Is Human*: An Assessment Of Progress And Emerging Priorities In Patient Safety. *Health Aff* 2018;37:1736–43.
- Singh H, Graber ML. Improving diagnosis in health care — the next imperative for patient safety. *N Engl J Med* 2015;373:2493–5.
- Newman-Toker DE, Moy E, Valente E, *et al*. Missed diagnosis of stroke in the emergency department: a cross-sectional analysis of a large population-based sample. *Diagnosis* 2014;1:155–66.
- El-Kareh R, Hasan O, Schiff GD. Use of health information technology to reduce diagnostic errors. *BMJ Qual Saf* 2013;22:ii40–51.
- Greenes RA, Bates DW, Kawamoto K, *et al*. Clinical decision support models and frameworks: seeking to address research issues underlying implementation successes and failures. *J Biomed Inform* 2018;78:134–43.
- Bates DW, Kuperman GJ, Wang S, *et al*. Ten commandments for effective clinical decision support: making the practice

- of evidence-based medicine a reality. *J Am Med Inform Assoc* 2003;10:523–30.
- 8 Karsh BT. *Clinical practice improvement and redesign: how change in workflow can be supported by clinical decision support*. AHRQ national resource center for health it update. Rockville, Maryland: Agency for Healthcare Research and Quality, 2009.
  - 9 Kilsdonk E, Peute LW, Jaspers MWM. Factors influencing implementation success of guideline-based clinical decision support systems: a systematic review and gaps analysis. *Int J Med Inform* 2017;98:56–64.
  - 10 Sittig DF, Wright A, Osheroff JA, *et al*. Grand challenges in clinical decision support. *J Biomed Inform* 2008;41:387–92.
  - 11 Ratwani RM, Reider J, Singh H. A decade of health information technology usability challenges and the path forward. *JAMA* 2019;321:743–4.
  - 12 Perry SJ, Wears RL, Fairbanks RJ. Human factors and ergonomics in the emergency department. In: Carayon P, ed. *Handbook of human factors and ergonomics in health care and patient safety*. 2nd edn. Boca Raton, FL: Taylor & Francis Group, 2012: 709–22.
  - 13 Hall KK, Schenkel SM, Hirshon JM, *et al*. Incidence and types of non-ideal care events in an emergency department. *Qual Saf Health Care* 2010;19(Suppl 3):i20–5.
  - 14 Schiff GD, Bates DW. Can electronic clinical documentation help prevent diagnostic errors? *N Engl J Med* 2010;362:1066–9.
  - 15 Mazur LM, Mosaly PR, Moore C, *et al*. Association of the usability of electronic health records with cognitive workload and performance levels among physicians. *JAMA Netw Open* 2019;2:e191709.
  - 16 Bennett P, Hardiker NR. The use of computerized clinical decision support systems in emergency care: a substantive review of the literature. *J Am Med Inform Assoc* 2017;24:655–68.
  - 17 Patterson BW, Pulia MS, Ravi S, *et al*. Scope and influence of electronic health Record-Integrated clinical decision support in the emergency department: a systematic review. *Ann Emerg Med* 2019;74:285–96.
  - 18 Deblois S, Chartrand-Lefebvre C, Toporowicz K, *et al*. Interventions to reduce the overuse of imaging for pulmonary embolism: a systematic review. *J Hosp Med* 2018;13:52–61.
  - 19 Dean NC, Jones BE, Jones JP, *et al*. Impact of an electronic clinical decision support tool for emergency department patients with pneumonia. *Ann Emerg Med* 2015;66:511–20.
  - 20 Jun S, Plint AC, Campbell SM, *et al*. Point-of-care cognitive support technology in emergency departments: a scoping review of technology acceptance by clinicians. *Acad Emerg Med* 2018;25:494–507.
  - 21 Drescher FS, Chandrika S, Weir ID, *et al*. Effectiveness and acceptability of a computerized decision support system using modified wells criteria for evaluation of suspected pulmonary embolism. *Ann Emerg Med* 2011;57:613–21.
  - 22 Raja AS, Ip IK, Prevedello LM, *et al*. Effect of computerized clinical decision support on the use and yield of CT pulmonary angiography in the emergency department. *Radiology* 2012;262:468–74.
  - 23 Berner ES. *Clinical decision support systems: state of the art*. Rockville, MD: Agency for Healthcare Research and Quality, 2009.
  - 24 Goyard C, Côté B, Looten V, *et al*. Determinants and prognostic implication of diagnostic delay in patients with a first episode of pulmonary embolism. *Thromb Res* 2018;171:190–8.
  - 25 Schiff GD, Hasan O, Kim S, *et al*. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med* 2009;169:1881–7.
  - 26 Kline JA. Diagnosis and exclusion of pulmonary embolism. *Thromb Res* 2018;163:207–20.
  - 27 Rohacek M, Buatsi J, Szucs-Farkas Z, *et al*. Ordering CT pulmonary angiography to exclude pulmonary embolism: defense versus evidence in the emergency room. *Intensive Care Med* 2012;38:1345–51.
  - 28 Kline JA, Shapiro NI, Jones AE, *et al*. Outcomes and radiation exposure of emergency department patients with chest pain and shortness of breath and ultralow pretest probability: a multicenter study. *Ann Emerg Med* 2014;63:281–8.
  - 29 Ouellette DW, Patocka C. Pulmonary embolism. *Emerg Med Clin North Am* 2012;30:329–75.
  - 30 Ageno W, Agnelli G, Imberti D, *et al*. Factors associated with the timing of diagnosis of venous thromboembolism: results from the master registry. *Thromb Res* 2008;121:751–6.
  - 31 Middleton B, Bloomrosen M, Dente MA, *et al*. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from Amia. *J Am Med Inform Assoc* 2013;20:e2–8.
  - 32 Miller K, Mosby D, Capan M, *et al*. Interface, information, interaction: a narrative review of design and functional requirements for clinical decision support. *J Am Med Inform Assoc* 2018;25:585–92.
  - 33 Miller K, Capan M, Weldon D, *et al*. The design of decisions: Matching clinical decision support recommendations to Nielsen's design heuristics. *Int J Med Inform* 2018;117:19–25.
  - 34 Schnittker R, Marshall SD, Horberry T, *et al*. Decision-centred design in healthcare: the process of identifying a decision support tool for airway management. *Appl Ergon* 2019;77:70–82.
  - 35 Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. *IEEE Trans Syst Man Cybern A* 2000;30:286–97.
  - 36 Scapin DL, Bastien JMC. Ergonomic criteria for evaluating the ergonomic quality of interactive systems. *Behav Inf Technol* 1997;16:220–31.
  - 37 Zhang J, Johnson TR, Patel VL, *et al*. Using usability heuristics to evaluate patient safety of medical devices. *J Biomed Inform* 2003;36:23–30.
  - 38 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 Inform Assoc* 2014;21:e287–96.
  - 39 Chan J, Shojanian KG, Easty AC, *et al*. Does user-centred design affect the efficiency, usability and safety of CPOE order sets? *J Am Med Inform Assoc* 2011;18:276–81.
  - 40 Militello L, Diulio J, Borders M, *et al*. Evaluating a modular decision support application for colorectal cancer screening. *Appl Clin Inform* 2017;26:162–79.
  - 41 Saleem JJ, Patterson ES, Militello L, *et al*. Impact of clinical reminder redesign on learnability, efficiency, usability, and workload for ambulatory clinic nurses. *J Am Med Inform Assoc* 2007;14:632–40.
  - 42 Zwaan L, Schiff GD, Singh H. Advancing the research agenda for diagnostic error reduction. *BMJ Qual Saf* 2013;22:ii52–7.
  - 43 Henriksen K, Brady J. The pursuit of better diagnostic performance: a human factors perspective. *BMJ Qual Saf* 2013;22:ii1–5.

- 44 Singh H, Graber ML, Kissam SM, *et al.* System-Related interventions to reduce diagnostic errors: a narrative review. *BMJ Qual Saf* 2012;21:160–70.
- 45 International Organization for Standardization. *ISO 9241-11 ergonomics of human-system interaction - part 11: usability: definitions and concepts*. Geneva, Switzerland: ISO, 2018.
- 46 Ammenwerth E, de Keizer NF. Publishing health IT evaluation studies. In: Ammenwerth E, Rigby M, eds. *Evidence-based health informatics*. IOS press, 2016: 304–11.
- 47 Talmon J, Ammenwerth E, Brender J, *et al.* STARE-HI—Statement on reporting of evaluation studies in Health Informatics. *Int J Med Inform* 2009;78:1–9.
- 48 Wells PS, Anderson DR, Rodger M, *et al.* Excluding pulmonary embolism at the bedside without diagnostic imaging: management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and D-dimer. *Ann Intern Med* 2001;135:98–107.
- 49 Raja AS, Greenberg JO, Qaseem A, *et al.* Evaluation of patients with suspected acute pulmonary embolism: best practice advice from the clinical guidelines Committee of the American College of physicians. *Ann Intern Med* 2015;163:701–11.
- 50 Kline JA, Peterson CE, Steuerwald MT. Prospective evaluation of real-time use of the pulmonary embolism Rule-out criteria in an academic emergency department. *Acad Emerg Med* 2010;17:1016–9.
- 51 Goldhaber SZ, Bounameaux H. Pulmonary embolism and deep vein thrombosis. *Lancet* 2012;379:1835–46.
- 52 Hoonakker PLT, Carayon P, Salwei M, *et al.* The Design of PE Dx, a CDS to Support Pulmonary Embolism Diagnosis in the ED. In: Marcilly R, Kuziemsky G, Nohr C, eds. *Context sensitive health informatics: sustainability in dynamic ecosystems*. Amsterdam, The Netherlands: IOS Press, 2019: 134–40.
- 53 Russ AL, Saleem JJ. Ten factors to consider when developing usability scenarios and tasks for health information technology. *J Biomed Inform* 2018;78:123–33.
- 54 Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock PA, Meshkati N, eds. *Human mental workload*. Amsterdam, The Netherlands: Elsevier Science Publishers, 1988: 139–83.
- 55 Lewis JR. Ibm computer usability satisfaction questionnaires: psychometric evaluation and Instructions for use. *Int J Hum Comput Interact* 1995;7:57–78.
- 56 Ozechowski TJ. Empirical Bayes MCMC estimation for modeling treatment processes, mechanisms of change, and clinical outcomes in small samples. *J Consult Clin Psychol* 2014;82:854–67.
- 57 Zhang Z, Hamagami F, Lijuan Wang L, *et al.* Bayesian analysis of longitudinal data using growth curve models. *Int J Behav Dev* 2007;31:374–83.
- 58 Lee SY. *Structural equation modeling. A Bayesian approach*. Chichester, UK: John Wiley & Sons, 2007.
- 59 Song XY, Lee SY. *Basic and advanced bayesian structural equation modeling: with applications in the medical and behavioral sciences*. Chichester, England: John Wiley & Sons, 2012.
- 60 Yanuar F. The estimation process in Bayesian structural equation modeling approach. *J Phys Conf Ser* 2014;495.
- 61 Salwei M, Carayon P, Hoonakker P, *et al.* Assessing workflow of emergency physicians in the use of clinical decision support. In: The Human Factors and Ergonomics Society, ed. *Proceedings of the annual conference of the human factors and ergonomics Society*. Washington, DC, 2019: 772–6.
- 62 Challenger DW, Prokop LJ, Abu-Saleh O. The proliferation of reports on clinical scoring systems: issues about uptake and clinical utility. *JAMA* 2019;321:2405–6.
- 63 Kroth PJ, Morioka-Douglas N, Veres S, *et al.* Association of electronic health record design and use factors with clinician stress and burnout. *JAMA Network Open* 2019;2:e199609–e09.
- 64 Ratwani RM, Savage E, Will A, *et al.* Identifying electronic health record usability and safety challenges in pediatric settings. *Health Aff* 2018;37:1752–9.
- 65 Carayon P, Wooldridge A, Hose B-Z, *et al.* Challenges and opportunities for improving patient safety through human factors and systems engineering. *Health Aff* 2018;37:1862–9.
- 66 Kushniruk A, Nohr C, Borycki E. Human factors for more usable and safer health information technology: where are we now and where do we go from here? *Yearb Med Inform* 2016:120–5.
- 67 Dalal AK, Fuller T, Garabedian P, *et al.* Systems engineering and human factors support of a system of novel EHR-integrated tools to prevent harm in the hospital. *J Am Med Inform Assoc* 2019;337.
- 68 Carayon P, Hoonakker P. Human factors and usability for health information technology: old and new challenges. *Yearb Med Inform* 2019;28:071–7.
- 69 Topol EJ. High-Performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.
- 70 Jamieson T, Goldfarb A. Clinical considerations when applying machine learning to decision-support tasks versus automation. *BMJ Qual Saf* 2019;28:778–81.
- 71 Celi LA, Fine B, Stone DJ. An awakening in medicine: the partnership of humanity and intelligent machines. *Lancet Digit Health* 2019;1:e255–7.
- 72 Carayon P, Schoofs Hundt A, Karsh B-T, *et al.* Work system design for patient safety: the SEIPS model. *Qual Saf Health Care* 2006;15(Suppl 1):i50–8.
- 73 Carayon P, Kianfar S, Li Y, *et al.* A systematic review of mixed methods research on human factors and ergonomics in health care. *Appl Ergon* 2015;51:291–321.
- 74 Sanders MS, McCormick EJ. *Human factors in engineering and design*. New York, NY: McGraw-Hill, 1993.
- 75 Becker BJ. Synthesizing standardized mean-change measures. *Br J Math Stat Psychol* 1988;41:257–78.