Original Investigation | Health Informatics

Association of the Usability of Electronic Health Records With Cognitive Workload and Performance Levels Among Physicians

Lukasz M. Mazur, PhD; Prithima R. Mosaly, PhD; Carlton Moore, MD; Lawrence Marks, MD

Abstract

IMPORTANCE Current electronic health record (EHR) user interfaces are suboptimally designed and may be associated with excess cognitive workload and poor performance.

OBJECTIVE To assess the association between the usability of an EHR system for the management of abnormal test results and physicians' cognitive workload and performance levels.

DESIGN, SETTING, AND PARTICIPANTS This quality improvement study was conducted in a simulated EHR environment. From April 1, 2016, to December 23, 2016, residents and fellows from a large academic institution were enrolled and allocated to use either a baseline EHR (n = 20) or an enhanced EHR (n = 18). Data analyses were conducted from January 9, 2017, to March 30, 2018.

INTERVENTIONS The EHR with enhanced usability segregated in a dedicated folder previously identified critical test results for patients who did not appear for a scheduled follow-up evaluation and provided policy-based decision support instructions for next steps. The baseline EHR displayed all patients with abnormal or critical test results in a general folder and provided no decision support instructions for next steps.

MAIN OUTCOMES AND MEASURES Cognitive workload was quantified subjectively using NASA–Task Load Index and physiologically using blink rates. Performance was quantified according to the percentage of appropriately managed abnormal test results.

RESULTS Of the 38 participants, 25 (66%) were female. The 20 participants allocated to the baseline EHR compared with the 18 allocated to the enhanced EHR demonstrated statistically significantly higher cognitive workload as quantified by blink rate (mean [SD] blinks per minute, 16 [9] vs 24 [7]; blink rate, –8 [95% CI, –13 to –2]; *P* = .01). The baseline group showed statistically significantly poorer performance compared with the enhanced group who appropriately managed 16% more abnormal test results (mean [SD] performance, 68% [19%] vs 98% [18%]; performance rate, –30% [95% CI, –40% to –20%]; *P* < .001).

CONCLUSIONS AND RELEVANCE Relatively basic usability enhancements to the EHR system appear to be associated with better physician cognitive workload and performance; this finding suggests that next-generation systems should strip away non–value-added EHR interactions, which may help physicians eliminate the need to develop their own suboptimal workflows.

JAMA Network Open. 2019;2(4):e191709. doi:10.1001/jamanetworkopen.2019.1709

Key Points

Question Is enhanced usability of an electronic health record system associated with physician cognitive workload and performance?

Findings In this quality improvement study, physicians allocated to perform tasks in an electronic health record system with enhancement demonstrated statistically significantly lower cognitive workload; those who used a system with enhanced longitudinal tracking appropriately managed statistically significantly more abnormal test results compared with physicians allocated to use the baseline electronic health record.

Meaning Usability improvements in electronic health records appear to be associated with improved cognitive workload and performance levels among clinicians; this finding suggests that next-generation systems should strip away non–value-added interactions.

Author affiliations and article information are listed at the end of this article.

Introduction

The usability of electronic health records (EHRs) continues to be a major concern.[1-3] Usability challenges include suboptimal design of interfaces that have confusing layouts and contain either too much or too little relevant information as well as workflows and alerts that are burdensome. Suboptimal usability has been associated with clinician burnout and patient safety events, and improving the usability of EHRs is an ongoing need.[4,5]

A long-standing challenge for the US health care system has been to acknowledge and appropriately manage abnormal test results and associated missed or delayed diagnoses.[6-11] The unintended consequences of these shortcomings include missed and delayed cancer diagnoses and associated negative clinical outcomes (eg, 28% of women did not receive timely follow-up for abnormal Papanicolaou test results[8]; 28% of women requiring immediate or short-term follow-up for abnormal mammograms did not receive timely follow-up care[9]). Even in the EHR environment, with alerts and reminders in place, physicians continue to often inappropriately manage abnormal test results.[12-21] Some key remaining barriers to effective management of test results are suboptimal usability of existing EHR interfaces and the high volume of abnormal test result alerts, especially less-critical alerts that produce clutter and distract from the important ones.[22,23] In addition, few organizations have explicit policies and decision support systems in their EHR systems for managing abnormal test results, and many physicians have developed processes on their own.[24-26] These issues are among the ongoing reasons to improve the usability of the EHR-based interfaces for the evaluation and management of abnormal test results.

We present the results of a quality improvement study to assess a relatively basic intervention to enhance the usability of an EHR system for the management of abnormal test results. We hypothesized that improvements in EHR usability would be associated with improvements in cognitive workload and performance among physicians.

Methods

Participants

This research was reviewed and approved by the institutional review board committee of the University of North Carolina at Chapel Hill. Written informed consent was obtained from all participants. The study was performed and reported according to the Standards for Quality Improvement Reporting Excellence (SQUIRE) guideline.[27]

Invitations to participate in the study were sent to all residents and fellows in the school of medicine at a large academic institution, clearly stating the need for experience with using the Epic EHR software (Epic Systems Corporation) in reviewing test results to undergo the study's simulated scenarios. A $100 gift card was offered as an incentive for participation. Potential participants were given an opportunity to review and sign a consent document, which included information on study purpose, goals, procedures, and risks and rewards as well as the voluntary nature of participation and the confidentiality of data. Recruited individuals had the right to discontinue participation at any time. Forty individuals were recruited to participate, 2 of whom were excluded (eg, numerous cancellations), leaving 38 evaluable participants (**Table 1**).

Study Design

From April 1, 2016, to December 23, 2016, 38 participants were enrolled and prospectively and blindly allocated to a simulated EHR environment: 20 were assigned to use a baseline EHR (without changes to the interface), and 18 were assigned to use enhanced EHRs (with changes intended to enhance longitudinal tracking of abnormal test results in the system) (**Figure**). Abnormalities requiring an action included new abnormal test results and previously identified abnormal test results for patients who did not show up (without cancellation) for their scheduled appointment in which the findings would be addressed. The new abnormal test results included a critically abnormal

mammogram (BI-RADS 4 and 5) and Papanicolaou test result with high-grade squamous intraepithelial lesion as well as noncritical results for rapid influenza test, streptococcal culture complete blood cell count, basic metabolic panel, and lipid profile, among others. The previously identified critical test results that required follow-up included abnormal mammogram (BI-RADS 4 and 5), Papanicolaou test result with high-grade squamous intraepithelial lesion, chest radiograph with 2 × 2-cm lesion in the left upper lobe, pulmonary function test result consistent with severe restrictive lung disease, and pathologic examination with biopsy finding of ascending colon consistent with adenocarcinoma.

The simulated scenarios were iteratively developed and tested by an experienced physician and human factors engineer (C.M. and L.M.) in collaboration with an Epic software developer from the participating institution. The process included functionality and usability testing and took approximately 12 weeks to complete. The experimental design was based on previous findings that attending physicians use the EHR to manage approximately 57 test results per day over multiple interactions.[22,23] Given that residents often manage a lower volume of patients, the present study was designed such that participants were asked to review a total of 35 test results, including 8 or 16 abnormal test results evenly distributed between study groups, in 1 test session. By organizational policies and procedures, participants were expected to review all results, acknowledge and follow-up on abnormal test results, and follow-up on patients with a no-show status (without cancellation) for their scheduled appointment aimed at addressing their previously identified abnormal test result.

Table 1. Composition of Participants

| Variable | No. (%) | | | | | |
	Internal Medicine Specialty	Family Medicine Specialty	Pediatrics Specialty	Surgery Specialty	Other Specialty	Total
All patients	14 (37)	4 (11)	9 (24)	5 (13)	6 (16)	38
Baseline EHR	9 (45)	3 (15)	3 (15)	2 (10)	3 (15)	20
Enhanced EHR	5 (28)	1 (6)	6 (33)	3 (17)	3 (17)	18
Postgraduate year						
1	4 (40)	1 (10)	3 (30)	1 (10)	1 (10)	10
2	2 (25)	1 (13)	2 (25)	2 (25)	1 (13)	8
3	5 (45)	1 (9)	4 (36)	0	1 (9)	11
4	3 (43)	1 (14)	0	1 (14)	2 (29)	7
5	0	0	0	1 (50)	1 (50)	2
Sex						
Male	5 (38)	2 (15)	2 (15)	1 (8)	3 (23)	13
Female	9 (43)	2 (4)	7 (28)	4 (16)	3 (12)	25

Abbreviation: EHR, electronic health record.

Figure. Study Design
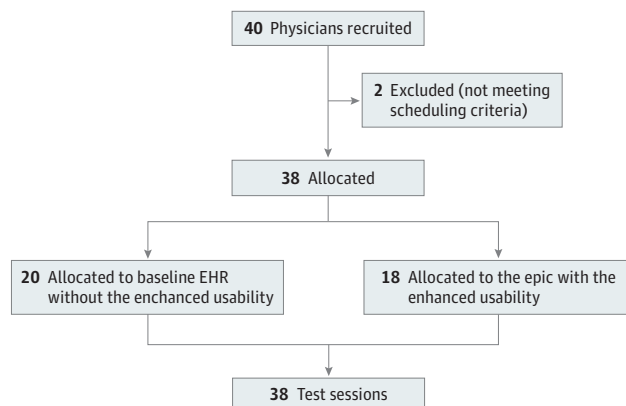
EHR indicates electronic health record.

The patient data in the simulation included full medical records, such as other clinicians' notes, previous tests, and other visits or subspecialist coverage.

Intervention

The baseline EHR (without enhanced interface usability), currently used at the study institution, displayed all new abnormal test results and previously identified critical test results for patients with a no-show status (did not show up for or cancelled their follow-up appointment) in a general folder called Results and had basic sorting capabilities. For example, it moved all abnormal test results with automatically flagged alerts to the top of the in-basket queue; flagged alerts were available only for test results with discrete values. Thus, critical test results for mammography, Papanicolaou test, chest radiograph, pulmonary function test, and pathologic examination were not flagged or sortable in the baseline EHR. The baseline EHR included patient status (eg, completed the follow-up appointment, no show), however, that information needed to be accessed by clicking on the visit or patient information tab located on available prebuilt views within each highlighted result.

The enhanced EHR (with enhanced interface usability) automatically sorted all previously identified critical test results for patients with a no-show status in a dedicated folder called All Reminders. It also clearly displayed information regarding patient status and policy-based decision support instructions for next steps (eg, "No show to follow-up appointment. Reschedule appointment in Breast Clinic").

The intervention was developed according to the classic theory of attention.[28] This theory indicates that cognitive workload varies continuously during the course of performing a task and that the changes of cognitive workload may be attributed to the adaptive interaction strategies of the operator exposed to task demands (eg, baseline or enhanced usability).

Main Outcomes and Measures

Perceived Workload

The NASA–Task Load Index (NASA-TLX) is a widely applied and valid tool used to measure workload,[29-34] including the following 6 dimensions: (1) mental demand (How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?); (2) physical demand (How much physical activity was required? Was the task easy or demanding, slack or strenuous?); (3) temporal demand (How much time pressure did you feel with regard to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?); (4) overall performance (How successful were you in performing the task? How satisfied were you with your performance?); (5) frustration level (How irritated, stressed, and annoyed [compared with content, relaxed, and complacent] did you feel during the task?); and (6) effort (How hard did you have to work, mentally and physically, to accomplish your level of performance?).

At the end of the test session, each participant performed 15 separate pairwise comparisons of the 6 dimensions (mental demand, physical demand, temporal demand, overall performance, frustration level, and effort) to determine the relevance (and hence weight) of a dimension for a given session for a participant. Next, participants marked a workload score between low (corresponding to 0) to high (corresponding to 100), separated by 5-point marks on the tool, for each dimension for each session. The composite NASA-TLX score for each session was obtained by multiplying the dimension weight with the corresponding dimension score, summing across all dimensions, and dividing by 15.

Physiological Workload

Using eye-tracking technology (Tobii X2-60 screen mount eye tracker; Tobii), we quantified physiological workload with validated methods based on changes in blink rate.[35,36] Eye closures ranging between 100 milliseconds to 400 milliseconds were coded as a blink. The validity (actual blink or loss of data) was later confirmed by visual inspection by the expert researcher on our team (P.R.M.) who specializes in physiological measures of cognitive workload. Decreased blink rate has

been found to occur in EHR-based tasks requiring more cognitive workload.[37] The fundamental idea is that blink rate slows down under visual task demands that require more focused attention and working memory load, but this association might vary with the type of visual task demands.[38-40] For each participant, the time-weighted mean blink rate measured during the participant's review of all abnormal test results was calculated and then considered for data analysis.

Performance

For each participant, performance was quantified as the percentage of (new or previously identified) abnormal test results that were appropriately acted on (with possible scores ranging from 0%-100%). Appropriate action on abnormal test result was defined as the study participant ordering (compared with not ordering) a referral for further diagnostic testing (eg, breast biopsy for mass identified on an abnormal mammogram) to a subspecialty clinic (eg, breast clinic). In addition, per the policy and procedures of the institution in which the study took place, if patients missed their appointment for follow-up on critical test results, the participants were expected to contact (compared with not contact) schedulers to reschedule follow-up care. We also quantified the total amount of time that participants took to complete each simulated scenario.

Secondary Outcome and Measure

Fatigue can affect perceived and physiological workload and performance and thus can confound study results.[41-43] Because of the possible confounding association of fatigue, participants were asked to evaluate their own state of fatigue immediately before each simulated session using the fatigue portion of the Crew Status Survey.[44] The fatigue assessment scale included these levels: 1 (fully alert, wide awake, or extremely peppy), 2 (very lively, or responsive but not at peak), 3 (okay, or somewhat fresh), 4 (a little tired, or less than fresh), 5 (moderately tired, or let down), 6 (extremely tired, or very difficult to concentrate), and 7 (completely exhausted, unable to function effectively, or ready to drop). The Crew Status Survey has been tested in real and simulated environments and has been found to be both reliable and able to discriminate between fatigue levels.[44,45]

Statistical Analysis

On the basis of the anticipated rate of appropriately identified abnormal test results in the literature[12-21] and the anticipated magnitude of the association of the enhanced EHR, we required a sample size of 30 participants, each reviewing 35 test results, to achieve 80% power to detect a statistically significant difference in cognitive workload and performance. Specifically, we performed sample size calculations at $\alpha = .05$, assuming that we could detect a mean (SD) difference of 10 (10) in NASA-TLX scores, a mean (SD) difference of 5 (10) in blink rate, and a mean (SD) difference of 10% (15%) in performance.

Before data analyses, we completed tests for normality using the Shapiro-Wilk test and equal variance using the Bartlett test for all study variables (cognitive workload, performance, and fatigue). Results indicated that all assumptions to perform parametric data analysis were satisfied (normality: all $P > .05$; equal variance: all $P > .05$).

We conducted a 2-sample t test to assess the association of enhanced usability of the EHR interface to manage abnormal test results with physician cognitive workload and performance. All data analyses were conducted from January 9, 2017, to March 30, 2018, using JMP 13 Pro software (SAS Institute Inc). Statistical significance level was set at 2-sided $P = .05$, with no missing data to report.

Results

Of the 852 eligible residents and fellows, 38 (5%) participated. Twenty-five participants (66%) were female and 13 (34%) were male. Thirty-six (95%) were residents and 2 (5%) were fellows (Table 1). Descriptive statistics of cognitive workload and performance are provided in **Table 2**.

Perceived and Physiological Workload

No statistically significant difference was noted in perceived workload between the baseline EHR and enhanced EHR groups (mean [SD] NASA-TLX score, 53 [14] vs 49 [16]; composite score, 4 [95% CI, –5 to 13]; *P* = .41). A statistically significantly higher cognitive workload as shown by the lower mean blink rate was found in the baseline EHR group compared with the enhanced EHR group (mean [SD] blinks per minute, 16 [9] vs 24 [7]; blink rate, –8 [95% CI, –13 to –2]; *P* = .01).

Performance

A statistically significantly poorer performance was found in the baseline EHR group compared with the enhanced EHR group (mean [SD] performance, 68% [19%] vs 98% [18%]; performance rate, –30% [95% CI, –40% to –20%]; *P* < .001). The difference was mostly attributable to review of patients with a no-show status for a follow-up appointment (Table 2). No difference between the baseline and enhanced EHR groups was noted in time to complete simulated scenarios (mean [SD] time in seconds, 238 [83] vs 236 [77]; time to complete, 2 seconds [95% CI, –49 to 52]; *P* > .05). No statistically significant difference was noted in fatigue levels between baseline and enhanced EHR groups (mean [SD] fatigue level, 2.7 [1.4] vs 2.8 [0.9]; fatigue level, –0.1 [95% CI, –0.8 to 0.7]; *P* = .84).

The rate of appropriately managing previously identified critical test results of patients with a no-show status in the baseline EHR was 37% (34 of 90 failure opportunities) compared with 77% (62 of 81 failure opportunities) in the enhanced EHR. The rate of appropriately acknowledging new abnormal test results in the baseline EHR group was 98% (118 of 120 failure opportunities; 2 participants did not acknowledge a critical Papanicolaou test result) compared with 100% (108 of 108 failure opportunities) in the enhanced EHR group.

Discussion

Participants in the enhanced EHR group indicated physiologically lower cognitive workload and improved clinical performance. The magnitude of the association of EHR usability with performance we found in the present study was modest, although many such improvements tend to have substantial value in the aggregate. Thus, meaningful usability changes can and should be implemented within EHRs to improve physicians' cognitive workload and performance. To our knowledge, this research is the first prospective quality improvement study of the association of EHR

Table 2. Perceived and Physiological Quantification of Cognitive Workload and Performance

Workload and Performance	Mean (SD)		
	Baseline EHR	Enhanced EHR	P Value
Perceived workload			
NASA-TLX score[a]	53 (14)	49 (16)	.41
Mental demand (mean weight: 3.67)	66 (15)	53 (19)	.02
Physical demand (mean weight: 0.19)	18 (10)	15 (12)	>.05
Temporal demand (mean weight: 2.83)	49 (24)	49 (22)	>.05
Performance demand (mean weight: 3.56)	37 (15)	39 (15)	>.05
Effort (mean weight: 2.67)	59 (21)	54 (17)	>.05
Frustration (mean weight: 2.08)	45 (28)	47 (21)	>.05
Cognitive workload			
Blink rate, blinks/min, No.	16 (9)	24 (7)	.01
Performance, No. appropriately managed/No. of failure opportunities (%)[b]			
Overall	152/210 (68)	170/189 (89)	<.001
New abnormal test results	118/120 (98)	108/108 (100)	>.05
Previously identified critical test results for patients with no-show status	34/90 (37)	62/81 (77)	<.001
Time to complete scenario, s	238 (83)	236 (77)	>.05

Abbreviations: EHR, electronic health record; NASA-TLX, NASA–Task Load Index.

[a] The NASA-TLX tool was used to measure workload,[29-34] including 6 dimensions. Score range: 0 (low) to 100 (high).

[b] Performance was the percentage of (new or previously identified) abnormal test results that were appropriately acted on. Possible scores ranged from 0% to 100%.

usability enhancements with both physiological measure of cognitive workload and performance during physicians' interactions with the test results management system in the EHR.

The enhanced EHR was more likely to result in participants reaching out to patients and schedulers to ensure appropriate follow-up. Physicians who used the baseline EHR were more likely to treat the EHR (not treat the patient) by duplicating the referral, rather than to reach out to patients and schedulers to find out the issues behind the no-show. In the poststudy conversations with participants, most indicated a lack of awareness about policies and procedures for managing patients with a no-show status and justified their duplication of orders as safer medial practice. This result seems to be in line with findings from real clinical settings, suggesting that few organizations have explicit policies and procedures for managing test results and most physicians developed processes on their own.[25,26]

The result from the baseline EHR group is in line with findings from real clinical settings that indicated physicians did not acknowledge abnormal test results in approximately 4% of cases.[19,20] The optimal performance in the enhanced EHR group is encouraging.

No significant difference was noted in the time to complete simulated scenarios and perceived workload between baseline and enhanced EHR groups, as quantified by the global NASA-TLX or by each dimension, while trending toward lower scores (Table 2). The time to complete simulated scenarios and NASA-TLX scores was elevated in the participants in the enhanced EHR group possibly because it was their first time interacting with this enhanced usability.

Overall, past and present research suggests that challenges remain in ensuring the appropriate management of abnormal test results. According to a study, 55% of clinicians believe that EHR systems do not have convenient usability for longitudinal tracking of and follow-up on abnormal test results, 54% do not receive adequate training on system functionality and usability, and 86% stay after hours or come in on the weekends to address notifications.[46]

We propose several interventions based on our findings to improve the proper management of abnormal test results. First, use the existing capabilities and usability features of the EHR interfaces to improve physicians' cognitive workload and performance. Similar recommendations were proposed by other researchers.[3,5,17-21,46-48] For example, the critical test results for patients with a no-show status should be flagged (ie, clearly visible to the clinician) indefinitely until properly acted on in accordance with explicit organizational policies and procedures. Second, develop explicit policies and procedures regarding the management of test results within EHRs, and implement them throughout the organization, rather than having clinicians develop their own approaches.[25,26,49] For example, Anthony et al[49] studied the implementation of a critical test results policy for radiology that defined critical results; categorized results by urgency and assigned appropriate timelines for communication; and defined escalation processes, modes of communication, and documentation processes. Measures were taken for 4 years from February 2006 to January 2010, and the percentage of reports adhering to the policies increased from 29% to 90%.[49] Third, given that the work is being done in an electronic environment, seize the opportunities to use innovative simulation-based training sessions to address the challenges of managing test results within an EHR ecosystem.[50-54] Fourth, establish a regular audit and feedback system to regularly give physicians information on their performance on managing abnormal test results.[55-57]

This study focused on a particular challenge (ie, the management of abnormal test results), but many other interfaces and workflows within EHRs can be similarly enhanced to improve cognitive workload and performance. For example, there is a need to improve reconciliation and management of medications, orders, and ancillary services. The next generation of EHRs should optimize usability by stripping away non–value-added EHR interactions, which may help eliminate the need for physicians to develop suboptimal workflows of their own.

Limitations

This study has several limitations, and thus caution should be exercised in generalizing the findings. First, the results are based on 1 experiment with 38 residents and fellows from a teaching hospital

artificially performing a discrete set of scenarios. Larger studies could consider possible confounding factors (eg, specialty, training levels, years of EHR use, attendings or residents) and more accurately quantify the association of usability with cognitive workload and performance. Second, performing the scenarios in the simulated environment, in which the participants knew that their work was going to be assessed, may have affected participants' performance (eg, more or less attentiveness and vigilance as perceived by being assessed or by the possibility of real harm to the patient). To minimize this outcome, all participants were given a chance to discontinue their participation at any time, but participant-specific findings would remain confidential. None of the participants discontinued participation in the study, although 2 participants were excluded from the study as they were not able to meet the scheduling criteria. Third, we acknowledge that the cognitive workload and performance scores were likely affected by the setting (eg, simulation laboratory and EHR) and thus might not reflect the actual cognitive workload and performance in real clinical settings. A laboratory setting cannot totally simulate the real clinical environment, and some activities cannot be easily reproduced (eg, looking up additional information about the patient using an alternative software, calling a nurse with a question about a particular patient, or a radiologist or laboratory technician calling physicians and verbally telling them about abnormal images). We also recognize that the enhanced usability was not optimal as it was designed and implemented within the existing capabilities of the EHR environment used for training purposes.

Fourth, the intervention might have manipulated both the ease of access to information through a reorganized display and learning because it provided a guide to action by clearly showing information on patient status and policy-based decision support instructions for next steps. Future research could more accurately quantify the association of usability and learning with cognitive workload and performance. Nevertheless, the intervention provided the necessary basis to conduct this study. All participants were informed about the limitations of the laboratory environment before the study began.

Conclusions

Relatively basic usability enhancements to EHR systems appear to be associated with improving physician management of abnormal test results while reducing cognitive workload. The findings from this study support the proactive evaluation of other similar usability enhancements that can be applied to other interfaces within EHRs.

Corresponding Author: Lukasz M. Mazur, PhD, Division of Healthcare Engineering, Department of Radiation Oncology, University of North Carolina at Chapel Hill, PO Box 7512, Chapel Hill, NC 27514 (lmazur@med.unc.edu).

Author Affiliations: School of Information and Library Science, University of North Carolina at Chapel Hill, Chapel Hill (Mazur, Mosaly, Moore); Carolina Health Informatics Program, University of North Carolina at Chapel Hill, Chapel Hill (Mazur, Mosaly, Moore); Division of Healthcare Engineering, Department of Radiation Oncology, University of North Carolina at Chapel Hill, Chapel Hill (Mazur, Mosaly, Marks); Division of General Medicine, University of North Carolina at Chapel Hill, Chapel Hill (Moore).

REFERENCES

1. Arndt BG, Beasley JW, Watkinson MD, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. *Ann Fam Med*. 2017;15(5):419-426. doi:10.1370/afm.2121

2. Middleton B, Bloomrosen M, Dente MA, et al; American Medical Informatics Association. 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(e1):e2-e8. doi:10.1136/amiajnl-2012-001458

3. Ratwani RM, Benda NC, Hettinger AZ, Fairbanks RJ. Electronic health record vendor adherence to usability certification requirements and testing standards. *JAMA*. 2015;314(10):1070-1071. doi:10.1001/jama.2015.8372

4. Shanafelt TD, Dyrbye LN, West CP. Addressing physician burnout: the way forward. *JAMA*. 2017;317(9): 901-902. doi:10.1001/jama.2017.0076

5. Howe JL, Adams KT, Hettinger AZ, Ratwani RM. Electronic health record usability issues and potential contribution to patient harm. *JAMA*. 2018;319(12):1276-1278. doi:10.1001/jama.2018.1171

6. McCarthy BD, Yood MU, Boohaker EA, Ward RE, Rebner M, Johnson CC. Inadequate follow-up of abnormal mammograms. *Am J Prev Med*. 1996;12(4):282-288. doi:10.1016/S0749-3797(18)30326-X

7. Peterson NB, Han J, Freund KM. Inadequate follow-up for abnormal Pap smears in an urban population. *J Natl Med Assoc*. 2003;95(9):825-832.

8. Yabroff KR, Washington KS, Leader A, Neilson E, Mandelblatt J. Is the promise of cancer-screening programs being compromised? quality of follow-up care after abnormal screening results. *Med Care Res Rev*. 2003;60(3): 294-331. doi:10.1177/1077558703254698

9. Jones BA, Dailey A, Calvocoressi L, et al. Inadequate follow-up of abnormal screening mammograms: findings from the race differences in screening mammography process study (United States). *Cancer Causes Control*. 2005;16(7):809-821. doi:10.1007/s10552-005-2905-7

10. Moore C, Saigh O, Trikha A, et al. Timely follow-up of abnormal outpatient test results: perceived barriers and impact on patient safety. *J Patient Saf*. 2008;4:241-244. doi:10.1097/PTS.0b013e31818d1ca4

11. Callen JL, Westbrook JI, Georgiou A, Li J. Failure to follow-up test results for ambulatory patients: a systematic review. *J Gen Intern Med*. 2012;27(10):1334-1348. doi:10.1007/s11606-011-1949-5

12. Kuperman GJ, Teich JM, Tanasijevic MJ, et al. Improving response to critical laboratory results with automation: results of a randomized controlled trial. *J Am Med Inform Assoc*. 1999;6(6):512-522. doi:10.1136/jamia.1999.0060512

13. Poon EG, Gandhi TK, Sequist TD, Murff HJ, Karson AS, Bates DW. "I wish I had seen this test result earlier!": dissatisfaction with test result management systems in primary care. *Arch Intern Med*. 2004;164(20):2223-2228. doi:10.1001/archinte.164.20.2223

14. Zapka J, Taplin SH, Price RA, Cranos C, Yabroff R. Factors in quality care–the case of follow-up to abnormal cancer screening tests–problems in the steps and interfaces of care. *J Natl Cancer Inst Monogr*. 2010;2010 (40):58-71. doi:10.1093/jncimonographs/lgq009

15. Lin JJ, Moore C. Impact of an electronic health record on follow-up time for markedly elevated serum potassium results. *Am J Med Qual*. 2011;26(4):308-314. doi:10.1177/1062860610385333

16. Laxmisan A, Sittig DF, Pietz K, Espadas D, Krishnan B, Singh H. Effectiveness of an electronic health record-based intervention to improve follow-up of abnormal pathology results: a retrospective record analysis. *Med Care*. 2012;50(10):898-904. doi:10.1097/MLR.0b013e31825f6619

17. Smith M, Murphy D, Laxmisan A, et al. Developing software to "track and catch" missed follow-up of abnormal test results in a complex sociotechnical environment. *Appl Clin Inform*. 2013;4(3):359-375. doi:10.4338/ACI-2013-04-RA-0019

18. Murphy DR, Meyer AND, Vaghani V, et al. Electronic triggers to identify delays in follow-up of mammography: harnessing the power of big data in health care. *J Am Coll Radiol*. 2018;15(2):287-295. doi:10.1016/j.jacr.2017.10.001

19. Singh H, Arora HS, Vij MS, Rao R, Khan MM, Petersen LA. Communication outcomes of critical imaging results in a computerized notification system. *J Am Med Inform Assoc*. 2007;14(4):459-466. doi:10.1197/jamia.M2280

20. Singh H, Thomas EJ, Mani S, et al. Timely follow-up of abnormal diagnostic imaging test results in an outpatient setting: are electronic medical records achieving their potential? *Arch Intern Med*. 2009;169(17):1578-1586. doi:10.1001/archinternmed.2009.263

21. Singh H, Thomas EJ, Sittig DF, et al. Notification of abnormal lab test results in an electronic medical record: do any safety concerns remain? *Am J Med*. 2010;123(3):238-244. doi:10.1016/j.amjmed.2009.07.027

22. Hysong SJ, Sawhney MK, Wilson L, et al. Provider management strategies of abnormal test result alerts: a cognitive task analysis. *J Am Med Inform Assoc*. 2010;17(1):71-77. doi:10.1197/jamia.M3200

23. Hysong SJ, Sawhney MK, Wilson L, et al. Understanding the management of electronic test result notifications in the outpatient setting. *BMC Med Inform Decis Mak*. 2011;11:22. doi:10.1186/1472-6947-11-22

24. Casalino LP, Dunham D, Chin MH, et al. Frequency of failure to inform patients of clinically significant outpatient test results. *Arch Intern Med*. 2009;169(12):1123-1129. doi:10.1001/archinternmed.2009.130

25. Elder NC, McEwen TR, Flach JM, Gallimore JJ. Management of test results in family medicine offices. *Ann Fam Med*. 2009;7(4):343-351. doi:10.1370/afm.961

26. Elder NC, McEwen TR, Flach J, Gallimore J, Pallerla H. The management of test results in primary care: does an electronic medical record make a difference? *Fam Med*. 2010;42(5):327-333.

27. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for Quality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf*. 2016;25(12):986-992. doi:10.1136/bmjqs-2015-004411

28. Kahneman D. *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall; 1973.

29. 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: North Holland Press; 1988:139-183. doi:10.1016/S0166-4115(08)62386-9

30. Ariza F, Kalra D, Potts HW. How do clinical information systems affect the cognitive demands of general practitioners? usability study with a focus on cognitive workload. *J Innov Health Inform*. 2015;22(4):379-390. doi:10.14236/jhi.v22i4.85

31. Mazur LM, Mosaly PR, Moore C, et al. Toward a better understanding of task demands, workload, and performance during physician-computer interactions. *J Am Med Inform Assoc*. 2016;23(6):1113-1120. doi:10.1093/jamia/ocw016

32. Young G, Zavelina L, Hooper V. Assessment of workload using NASA Task Load Index in perianesthesia nursing. *J Perianesth Nurs*. 2008;23(2):102-110. doi:10.1016/j.jopan.2008.01.008

33. Yurko YY, Scerbo MW, Prabhu AS, Acker CE, Stefanidis D. Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *Simul Healthc*. 2010;5(5):267-271. doi:10.1097/SIH.0b013e3181e3f329

34. Mazur LM, Mosaly PR, Jackson M, et al. Quantitative assessment of workload and stressors in clinical radiation oncology. *Int J Radiat Oncol Biol Phys*. 2012;83(5):e571-e576. doi:10.1016/j.ijrobp.2012.01.063

35. Beatty J, Lucero-Wagoner B. The pupillary system. In Cacioppo JT, Tassinary LG, Berston GG, eds. *Handbook of Psychophysiology*. New York, NY: Cambridge University Press; 2000:142-162.

36. Asan O, Yang Y. Using eye trackers for usability evaluation of health information technology: a systematic literature review. *JMIR Hum Factors*. 2015;2(1):e5. doi:10.2196/humanfactors.4062

37. Mosaly P, Mazur LM, Fei Y, et al. Relating task demand, mental effort and task difficulty with physicians' performance during interactions with electronic health records (EHRs). *Int J Hum Comput Interact*. 2018;34:467-475. doi:10.1080/10447318.2017.1365459

38. Fukuda K. Analysis of eyeblink activity during discriminative tasks. *Percept Mot Skills*. 1994;79(3 Pt 2): 1599-1608. doi:10.2466/pms.1994.79.3f.1599

39. Siyuan C, Epps J. Using task-induced pupil diameter and blink rate to infer cognitive load. *Hum Comput Interact*. 2014;29(4):390-413. doi:10.1080/07370024.2014.892428

40. Ueda Y, Tominaga A, Kajimura S, Nomura M. Spontaneous eye blinks during creative task correlate with divergent processing. *Psychol Res*. 2016;80(4):652-659. doi:10.1007/s00426-015-0665-x

41. Needleman J, Buerhaus P, Pankratz VS, Leibson CL, Stevens SR, Harris M. Nurse staffing and inpatient hospital mortality. *N Engl J Med*. 2011;364(11):1037-1045. doi:10.1056/NEJMsa1001025

42. van den Hombergh P, Künzi B, Elwyn G, et al. High workload and job stress are associated with lower practice performance in general practice: an observational study in 239 general practices in the Netherlands. *BMC Health Serv Res*. 2009;9:118. doi:10.1186/1472-6963-9-118

43. Weigl M, Müller A, Vincent C, Angerer P, Sevdalis N. The association of workflow interruptions and hospital doctors' workload: a prospective observational study. *BMJ Qual Saf*. 2012;21(5):399-407. doi:10.1136/bmjqs-2011-000188

44. Miller JC, Narveaz AA. A comparison of the two subjective fatigue checklists. *Proceedings of the 10th Psychology in the DoD Symposium*. Colorado Springs, CO: United States Air Force Academy; 1986:514-518.

45. Gawron VJ. *Human Performance, Workload, and Situational Awareness Measurement Handbook*. Boca Raton, FL: CRC Press; 2008.

46. Singh H, Spitzmueller C, Petersen NJ, et al. Primary care practitioners' views on test result management in EHR-enabled health systems: a national survey. *J Am Med Inform Assoc*. 2013;20(4):727-735. doi:10.1136/amiajnl-2012-001267

47. Ratwani RM, Savage E, Will A, et al. Identifying electronic health record usability and safety challenges in pediatric settings. *Health Aff (Millwood)*. 2018;37(11):1752-1759. doi:10.1377/hlthaff.2018.0699

48. Savage EL, Fairbanks RJ, Ratwani RM. Are informed policies in place to promote safe and usable EHRs? a cross-industry comparison. *J Am Med Inform Assoc*. 2017;24(4):769-775. doi:10.1093/jamia/ocw185

49. Anthony SG, Prevedello LM, Damiano MM, et al. Impact of a 4-year quality improvement initiative to improve communication of critical imaging test results. *Radiology*. 2011;259(3):802-807. doi:10.1148/radiol.11101396

50. Steadman RH, Coates WC, Huang YM, et al. Simulation-based training is superior to problem-based learning for the acquisition of critical assessment and management skills. *Crit Care Med*. 2006;34(1):151-157.

51. Mazur LM, Mosaly PR, Tracton G, et al. Improving radiation oncology providers' workload and performance: can simulation-based training help? *Pract Radiat Oncol*. 2017;7(5):e309-e316. doi:10.1016/j.prro.2017.02.005

52. Mohan V, Scholl G, Gold JA. Intelligent simulation model to facilitate EHR training. *AMIA Annu Symp Proc*. 2015;2015:925-932.

53. Milano CE, Hardman JA, Plesiu A, Rdesinski RE, Biagioli FE. Simulated electronic health record (Sim-EHR) curriculum: teaching EHR skills and use of the EHR for disease management and prevention. *Acad Med*. 2014;89 (3):399-403. doi:10.1097/ACM.0000000000000149

54. Stephenson LS, Gorsuch A, Hersh WR, Mohan V, Gold JA. Participation in EHR based simulation improves recognition of patient safety issues. *BMC Med Educ*. 2014;14:224. doi:10.1186/1472-6920-14-224

55. Weiner JP, Fowles JB, Chan KS. New paradigms for measuring clinical performance using electronic health records. *Int J Qual Health Care*. 2012;24(3):200-205. doi:10.1093/intqhc/mzs011

56. Rich WL III, Chiang MF, Lum F, Hancock R, Parke DW II. Performance rates measured in the American Academy of Ophthalmology IRIS Registry (Intelligent Research in Sight). *Ophthalmology*. 2018;125(5):782-784.

57. Austin JM, Demski R, Callender T, et al. From board to bedside: how the application of financial structures to safety and quality can drive accountability in a large health care system. *Jt Comm J Qual Patient Saf*. 2017;43(4): 166-175. doi:10.1016/j.jcjq.2017.01.001