International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: http://www.tandfonline.com/loi/hihc20

Relating Task Demand, Mental Effort and Task Difficulty with Physicians' Performance during Interactions with Electronic Health Records (EHRs)

Prithima Reddy Mosaly, Lukasz M. Mazur, Fei Yu, Hua Guo, Merck Derek, David H. Laidlaw, Carlton Moore, Lawrence B. Marks & Javed Mostafa

To cite this article: Prithima Reddy Mosaly, Lukasz M. Mazur, Fei Yu, Hua Guo, Merck Derek, David H. Laidlaw, Carlton Moore, Lawrence B. Marks & Javed Mostafa (2018) Relating Task Demand, Mental Effort and Task Difficulty with Physicians' Performance during Interactions with Electronic Health Records (EHRs), International Journal of Human–Computer Interaction, 34:5, 467-475, DOI: 10.1080/10447318.2017.1365459

To link to this article: https://doi.org/10.1080/10447318.2017.1365459

Accepted author version posted online: 24 Aug 2017.
Published online: 25 Sep 2017.

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
2018, VOL. 34, NO. 5, 467–475
https://doi.org/10.1080/10447318.2017.1365459

Taylor & Francis
Taylor & Francis Group

Check for updates

Relating Task Demand, Mental Effort and Task Difficulty with Physicians' Performance during Interactions with Electronic Health Records (EHRs)

Prithima Reddy Mosaly[a,b], Lukasz M. Mazur[a,b], Fei Yu[b], Hua Guo[c], Merck Derek[c], David H. Laidlaw[c], Carlton Moore[d], Lawrence B. Marks[a], and Javed Mostafa[b]

[a]Division of Healthcare Engineering, Department of Radiation Oncology, University of North Carolina Chapel Hill, Chapel Hill, NC, USA; [b]Carolina Health Informatics Program, School of Information and Library Science, University of North Carolina, Chapel Hill, NC, USA; [c]Department of Computer Science, Brown University, Providence RI, USA; [d]Division of General Medicine, University of North Carolina, Chapel Hill, NC, USA

ABSTRACT
Objective was to assess the relationship between task demand, mental effort, task difficulty, and performance during physicians' interaction with electronic health records (EHRs). Seventeen physicians performed three EHR-based scenarios with varying task demands. Mental effort was measured using eye tracking measures via task evoked pupillary responses (TEPR), blink frequency, and gaze speed; task difficulty (or user behavior) was measured using frequent mouse click patterns and task flow; user performance was quantified using two types of omission errors: (i) omission errors with *no* evidence of trying to complete the task and (ii) omission error with evidence of trying but unable to complete the task. The results indicated that task demand significantly increased mental effort, but not task difficulty. Task demand, mental effort, and task difficulty all predicted performance. Specifically, there was a significant relationship between (i) task demand, TEPR and omission errors with *no* evidence of trying to complete the task, and (ii) blink frequency, repeated search clicks and omission error with evidence of trying but unable to complete the task. In concert, results suggest that physicians' performance during EHR interaction was negatively affected by task demands and increase in mental effort. This highlight the need for implementation of appropriate quality assurance (QA) measures, in addition to EHR usability improvement, to minimize omission errors and improve physician's performance. Additionally, the lack of relationship between task demand and task difficulty highlights a need for further methodological and empirical studies to advance our understanding from theory to application during physician–EHR interaction.

1. Introduction

In healthcare, electronic health records (EHRs) offer the potential to improve patient care and decrease cost; however, serious unintended consequences from the implementation of these systems have emerged (Bowman, 2013). Poor usability of EHR design can "create new hazards in the already complex delivery of health care" (Lin & Stead, 2009), by increasing task demands, mental effort, and task difficulty that in turn negatively affects healthcare professionals' performance and patient safety (Elliott, Young, Brice, Aguiar & Kolm, 2014; Peute & Jaspers, 2007; Russ et al., 2014; Vicente, 1999). For example, Hill, Sears, and Melanson (2013) found that providers seeing (on average per hour) 2.4 patients require about 4,000 mouse-clicks in EHRs during a 10-hour shift. Further, more than 1,000 adverse events associated with EHRs were reported to the Pennsylvania Patient Safety Authority in 2011 (Sparnon & Marella, 2012). Recent reports that focused on EHR-related medical malpractice identified over 80% of the reported events involve patient harm: many deaths, strokes, missed and significantly delayed cancer diagnoses, massive hemorrhage, 10-fold overdoses, ignored or lost critical lab results, etc. (Graber, Siegal, Riah, Johnston, & Kenyon, 2015).

Overall, the relationship between task demands, mental effort, task difficulty (as experienced by the user due to interaction between user and task), and performance, has been vigorously studied in many domains including human–computer interaction (Hancock & Szalma, 2003; Merat, Antilla, & Luoma, 2005; Szalma & Hancock, 2007), but not much work have been done in physician–EHR interaction arena. Specific task (e.g., information structure and information rate), individual (e.g., cognitive capabilities, knowledge, experience), and technology (e.g., usability, functionality) characteristics were related to mental effort (perceived and physiological), behavior (e.g., strategies used to perform the task), and performance (Fairclough, Venables, & Tattersall, 2005; Goodhue & Thompson, 1995; Hancock, Williams, & Manning, 1995; Merat et al., 2005; Szalma & Hancock, 2007; Zhang & Galletta, 2006). For example, in air-traffic control (ATC) setting, increased rate of information (i.e., increased number of aircrafts under control) significantly increased mental effort and degraded performance (Chen, Epps, Ruiz, & Chen, 2011;

CONTACT Prithima Reddy Mosaly prithima_mosaly@med.unc.edu; prithimareddy@gmail.com Department of Radiation Oncology, University of North Carolina, Box 7512, Chapel Hill, NC 27514, USA.

Edward et al., 2017). In web search setting, Brennan, Kelly, Arguello, (2014) indicated that individuals with high cognitive capabilities experienced less perceived effort and improved search strategy. In healthcare setting, technological factors like poor usability increased experienced task demand and perceived effort (workload) leading to degraded performance (Mazur et al., 2016). Many of the above discussed learnings could be applied to improve physician–EHR interaction in healthcare setting. Therefore, there is an opportunity and need to quantify and better understand physicians' interaction with EHRs.

1.1. Theoretical foundation

According to limited resource theory, increase in task demands increases mental effort necessary for task performance and thus falters or fails performance (Kahneman, 1973; Moray, 1967). Specifically, the theory suggests that the effect on performance is both due to mental effort and task difficulty (i.e., how the task is experienced) and it is dependent on the context, state, capacity, and strategy of allocation of mental resources (De Waard, 1996; Kantowitz, 1987; Parasuraman & Hancock, 2001), which seems relevant during physicians' interaction with EHRs.

The concept of mental effort and task difficulty has been broadly studied to assess the effect of task demand on performance in other domains. This was done by associating task difficulty with operator's adaptive strategies (or behaviors) of resource allocation that enable them to manage overall workload and regulate their performance (Merat et al., 2005; Moray, Dessouky, Kijowski, & Adapathya, 1991; Parasuraman & Hancock, 2001; Szalma, 2002). Task demand is determined by the goal that must be attained by means of task performance and it is independent of the individual (De Waard, 1996; Parasuraman & Hancock, 2001). Mental effort is defined as the amount of cognitive resources supplied to perform a task and is reflected in manifestation of physiological arousal such as pupillary dilations (Kahneman, 1973). Task difficulty is the regulated behavior or adaptable strategies used to cope with increased task demand to mitigate mental effort required to perform the task (Kantowitz, 1987; Parasuraman & Hancock, 2001).

For example, under suboptimal driving conditions, drivers adapted task difficulty by reducing driving speed as a behavior strategy to optimize their mental effort and improve performance (Da Silva, 2014; De Waard, 1996). In aviation, Boehm-Davis et al. (2007) found that under high task demands (e.g., steep descent angles and low visibility conditions) and high perceived workload, pilots managed to safely land their aircraft by using a creative strategy of speed control and glide path in order to gain time to control the landing. Similar findings were drawn from the ATC (Athènes, Averty, Puechmorel, Delahaye, & Collet, 2002; Hilburn, 2004; Histon & Hansman, 2002; Loft, Sanderson, Neal, & MooiJ, 2007; Sperandio, 1971) and human–computer interaction domains (Hancock & Szalma, 2003; Hertzum & Holmegaard, 2013, Szalma & Hancock, 2007).

Given the raising number of errors within the EHR ecosystem, and relatively limited work done in the healthcare domain, it seems prudent that we quantify task difficulty and assess the relationship between task demand, mental effort, task difficulty, and performance during physicians' interaction with EHRs (Figure 1). Results from our study may help guide practitioners, users, researchers, developers, and vendors to choose/design proper EHR systems and improve training protocols to enhance patient safety in rapidly evolving EHR ecosystem.

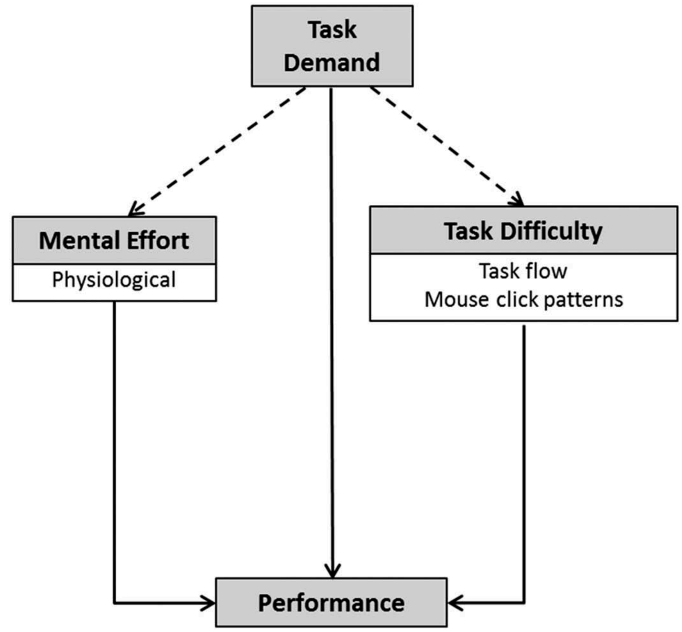

Figure 1. A conceptual model representing the relationship between task demand, mental effort, task difficulty and performance.
Dotted lines indicate univariate analysis to test the effect of task demand on mental effort and task difficulty respectively. The solid lines indicate multi-variable relationship testing between task demand, mental effort, task difficulty and performance.

2. Methods

2.1. Study and participants

The study was conducted within the Human Factors (HF) laboratory in the department of radiation oncology at University of North Carolina, Chapel Hill, where we have created a simulated environment that closely mirrors the real clinical environment of physician–EHR interactions. The Institutional Review Board (IRB) approved the study. Seventeen resident physicians and medical students' (6 post graduate year (PGY)-3+, 4 PGY-1, and 7 final year medical students) participated in the study. Eye movements and screen activities (computer mouse clicks) were monitored and captured using Tobii X-60 (60 Hz frequency) and Eyeworks©, while participants performed one baseline and three clinical scenarios.

2.2. Procedure

A brief introduction to the experiment/procedure, including eye tracking, was provided to the participants. First, to gauge participants' physiological baseline and responses, a basic working memory task of 3-letter memorization-recall (repeated 10 times) was administered. Next, three stimulated clinical scenarios involving test patients with (urinary tract

Table 1. Description of three clinical scenarios and list of instructed tasks for each scenario.

Scenario	Instructed tasks to be performed
(1) Urinary tract infection (UTI)	(1) Review the clinical and physical exam notes (2) Specify low risk for venous thromboembolism (VTE) prophylaxis *Note*: Ambulatory, out-of-bed and Education are sufficient prophylaxis (3) Order any necessary urine test(s) (4) Check results of test(s) (5) Order the appropriate treatment for the patient
(2) Pneumonia (PN)	(1) Review the clinical and physical exam notes (2) Specify low risk for venous thromboembolism (VTE) prophylaxis *Note*: Ambulatory, out-of-bed and Education are sufficient prophylaxis (3) Write admission orders: - Admit to Med wing "G" - Supplemental O_2/nasal cannula, wean per nursing - IV antibiotics; arterial blood gas (ABG); blood and sputum cultures; Posterior-anterior and Lateral chest X-ray (PA/Lat CXR) - AM Labs; Complete blood count with differential, basic metabolic pane (BMP) (5) Check results of tests and PA Lat CXR (6) Order Computed Tomography (CT) of chest w/o contrast (7) Check results of CT (8) Change to oral antibiotics and write discharge order; schedule for follow-up to Medicine clinic in 1 week
(3) Heart failure (HF)	(1) Review the History and Physical. This contains pertinent clinical history as well as physical exam (2) Specify high risk for venous thromboembolism (VTE) prophylaxis (3) Sub cutaneous heparin (4) Order labs: complete blood count (CBC), Chemistry, trans thoracic echocardiogram (TTE) (5) Check results of labs (6) Write admission orders including daily weights and low salt diet (7) Restart meds (8) Tobacco cessation consult (9) Order intravenous (IV) Lasix

infection [UTI], pneumonia [PN], and heart failure [HF]) were presented in the Epic© EHR playground environment (used at University of North Carolina at Chapel Hill). A 2-minute break was provided after the baseline and between the three clinical scenarios. Printed instructions describing pre-specified tasks on three routine clinical scenarios were provided next to the participant display monitor (Table 1).

2.3. Measures and assessment

Data collection

Task demand. The three clinical scenarios involved EHR task management for patients with UTI, PN, and HF, having variable range of tasks demands were presented in the same sequence (Mazur et al., 2016). These medical conditions were chosen because they represents some of the most common admission diagnoses to U.S. hospitals (Pfuntner, Wier, & Stocks, 2013) and therefore their management should be familiar to most practicing physicians. The UTI scenario was relatively straightforward and task demands were minimal as it required fewer decisions (4), followed by the HF (10) and PN (14) scenarios, both requiring more clinical decision making to design care for patients. Per experimental design and medical experts' opinion, task demand was considered as nominal independent variable with 3 levels (UTI, PN, and HF); and per analysis of computer mouse clicks and perceived workload (Mazur et al., 2016) task demand was also considered as nominal independent variable with 2 levels (UTI, PN, & HF). Thus, data analysis was conducted with task demands set at 3- and 2-levels respectively.

Mental effort. Mental effort was quantified physiologically using pupillary data based on task evoked pupillary response (TEPR), blink frequency, and gaze velocity using recommended procedures (Beatty & Lucero-Wagoner, 2000; Poole & Ball, 2006). The pupillary data was averaged to 1 sample/sec for all the experiments. A baseline pupil size was computed based on the basic 3-letter memorization-recall experiment (repeated 10 times), and consisting of 1 second pupillary data prior to the recall of the memorized letters (averaged across 10 trials). TEPR was calculated by subtracting the baseline pupil size from clinical task pupil size. Blink frequency was calculated by dividing the total number of blinks in a given task by the task time. Gaze velocity was computed as the distance (measured in degrees) of gaze travelled between two consecutive samples (data collected at 60 samples per second) divided by the inter-sample time, as recommended by Holmqvist et al. (2011). Mental effort was considered as continuous variable in data analysis.

Task difficulty. Task difficulty was quantified based on participant's behavior (strategic approach) to perform the task, i.e., by identifying and counting repeated patterns of mouse clicks and task flow analysis. Two researchers independently watched the recorded videos of each scenario captured using Eyeworks© software and coded mouse clicks into following three categories: (1) navigation clicks (e.g., moving from one tab to another for navigational purposes); (2) decision clicks (e.g., selecting a test or medication to order or cancelling a selected order or search result); and (3) Input clicks (e.g., placing the mouse cursor into the search box to type search terms). Any discrepancies (e.g., in some instances, navigation, and decision click codes were swapped) were resolved during our weekly meetings. Commonly occurring click patterns were extracted from the sequences using two-pass approach adapted from Guo, Gomez, Ziemkiewicz, and Laidlaw (2016). The first pass finds the most frequently occurred and possibly overlapping patterns, while the second pass segments each sequence into a series of patterns found in the first pass or singleton actions. The algorithm merges patterns with clicks appearing in the same order but for different number of time windows during the task. For example, a stream of three decision clicks is considered the same as a stream of four decision clicks, and their occurrences are combined when ranking patterns by frequency during the first pass. Patterns that are mutually exclusive (i.e., end of one pattern not the same as the beginning of another pattern) were considered for analysis. The patterns included in the study were (1) *editing order details* that involved a decision click (to select the order), multiple input clicks (to provide details like, dose, start date, frequency, comments, reason for order, etc.), and a decision click to accept the edits (represented as DI^mD); (2) *making orders* that involved multiple navigation clicks while reviewing notes or admission orders and then navigating to place (or select) multiple orders (represented as N^mD^m); (3) *reviewing/exploration of orders* that involved navigating to orders to review placed/selected orders or explore hidden sub-orders by expanding the listed orders (represented as NDN^m); (4) *searching for orders* that involved navigating to the search page, then searching for an order using keyword(s), followed by multiple selection (or cancellation of returned search results followed by a new keyword search) of the orders (represented as (NID^m).

Task difficulty was also quantified using task flow, measured as number of deviation from the instructed task sequence as presented in Table 1. For example (e.g., UTI scenario), if participant revisited (repeated) task 2 after performing (or during performance of) task 3, the deviation was counted as '1'. The task flow was quantified by counting total number of repeated tasks obtained via video analysis. Task difficulty was considered as nominal variable in data analysis.

Performance. Performance was quantified using two types of omission errors, (1) omission error with *no* evidence of trying to complete the task, and (ii) omission error with evidence of trying but unable to complete the task (e.g., tried but failed to complete the task like could not find medication or test; incorrectly/partially performed the tasks like ordering incorrect/insufficient medication or test). Two researchers evaluated performance and categorization of errors via video analysis. This was further reviewed and confirmed by an expert physician using the registry of completed orders and video recordings. If participant made both types of omission errors in a given scenario, their performance was coded as 'error' accordingly in each of the categories for relationship testing. Performance was considered as nominal dichotomous variable (no error versus omission error with no evidence of trying to complete the task; and no error versus omission error with evidence of trying but unable to complete the task) in data analysis.

Data analysis. Before data analysis, we completed tests for normality and equal variance for all study variables using Shapiro-Wilk's and Bartlett test respectively. Results indicated that assumptions were satisfied (normality: all $p > 0.05$; equal variance: all $p > 0.05$).

A series of univariate least square (or nominal logistic) regression analyses were performed to assess the relationship between task demand (3- and 2-levels), mental effort, and task difficulty, as represented in Figure 1. Task demand was considered as independent variable, whereas mental effort and task difficulty were considered as dependent measures. Participants were included in the analysis and considered as random factor. A post-hoc analysis using Tukey's HSD was performed to assess the significance between the three levels of task demand. Next, all variables (task demand, mental effort, and task difficulty) were included in a multivariable binomial logistic regression analysis to assess the relationship with performance. An alpha level of 0.05 was set for significance testing. All analyses were performed using JMP© 13 software and SPSS version 23©.

3. Results

Descriptive statistics of mental effort, task difficulty, and performance for three clinical scenarios (UTI, PN, HF) are provided in Table 2.

3.1. Task demand

Relationship with mental effort

There was a significant negative effect of task demands on blink frequency (3-levels: $F(2,29) = 7.4$, $p = 0.003$; 2-levels: $F(1,29) = 15$, $p < 0.001$), indicating that blink frequency was

Table 2. Descriptive statistics of Mental Effort, Task and Performance for the three clinical scenarios.

Measures		UTI	PN	HF
Mental effort	TEPR in mm[1] [mean(sd)]	0.15 (0.14)	0.16 (0.17)	0.14 (0.16)
	Blink Freq. as #/min [1] [mean(sd)]	8.7 (3)	4.3 (2)	4.2 (3)
	Gaze Speed in degrees/sec [1] [mean(sd)]	138 (46)	134 (29)	139 (43)
Task difficulty	Editing order details (Patter: DI^mD) [count]	2	6	3
	Making orders (Pattern: N^mD^m) [count]	32	64	36
	Reviewing/exploration of orders (Pattern: NDN^m) [count]	14	20	14
	Searching for orders (Pattern: NID^m) [count]	8	19	12
	Task flow [count]	8	13	10
Performance	Omission error with no evidence of trying to complete the task [count]	1	8	2
	Omission error with evidence of trying but unable to complete the task [count]	2	8	7

[1]The eye data on one participant was discarded due to excessive loss of data (>40%) for the three clinical scenarios.
N: navigation clicks; D: Decision clicks; I: Input clicks; N^m: two or more consecutive navigation clicks; D^m: two or more consecutive decision clicks; I^m: two or more consecutive input clicks.

significantly lower in PN and HF compared to UTI scenario ($p < 0.05$).

Relationship with task difficulty

There was no significant effect of task demand (both 3- and 2-levels) on task difficulty ($p > 0.05$).

3.2. Performance

Relationship with task demands, effort, and task difficulty

There was a significant positive relationship between omission errors with no evidence of trying to complete the task and task demand, and mental effort measured by TEPR (with 3-level task demand: $\chi^2_{(9,n=51)} = 25$, $p = 0.003$; with 2-level task demand: $\chi^2_{(8,n=51)} = 14$, $p = 0.05$). The odds of making an error in PN case was 15 and 10 times more likely compared to HF and UTI cases respectively for 3-levels of task demand; whereas the odds of making an error in PH & HF case grouped together was 16 times more likely compared to UTI case for 2-levels of task demand. Also, TEPR was 110% more dilated in participants who made an error compared to participants with no error ($p = 0.002$) (Table 3).

There was a significant negative relationship between omission errors with evidence of trying but unable to complete the task and mental effort measured by blink frequency and a significant positive relationship with click pattern, specifically with searching for orders (NID^m) (with 3-level task demand: $\chi^2_{(9,n=51)} = 21$, $p = 0.01$; with 2-level task demand: $\chi^2_{(8,n=51)} = 19$, $p = 0.01$). The odds of making an error by participants who performed repeated searching for orders (NID^m) was 19 (when task demand was included at 3-levels) and 17 (when task demand was included as 2-levels) times more likely compared to participants with no error ($p = 0.03$). Blink frequency was 50% lower in participant who made an error compared to the participants with no error ($p < 0.001$) (Table 3).

4. Discussion

There are number of key contributions of our study to the body of knowledge of human–computer interaction, especially to healthcare domain. First, we found that EHR-related task demands significantly increases mental effort, which is in line with previous findings including aviation (Ayaz et al., 2010; Colle & Reid, 2005; De Rivecourt, Kuperus, Post, & Mulder, 2008; Hancock et al., 1995; Hoffman, Pene, Rognin, & Zeghal, 2003; Lee, Kerns, Bone, & Nickelson, 2001), nuclear power (Byun & Choi, 2002; Liang et al., 2009), human–computer interaction (Cullen, Dan, Rogers, & Fisk, 2014; Hertzum, & Holmegaard, 2013), and in healthcare (Carswell, Clarke, & Seales, 2005; Mazur et al., 2016; Young, Zavelina, & Hooper, 2008; Yurko, Scerbo, Prabhu, Acker, & Stefanidis, 2010; Zhang, Padman, & Levin, 2012). Specifically, we found blink frequency to be a predictor of mental effort as evoked by task demands. Similar findings have been reported by other studies in various domains, commonly indicating that a lower blink frequency is assumed to indicate a higher visual workload (Brookings, Wilson, & Swain, 1996; Bruneau, Sasse, & McCarthy, 2002; May, Kennedy, Williams, Dunlap, & Brannan, 1990; Mosaly, Mazur, & Marks, 2016; Poole & Ball, 2006; Zheng et al., 2012).

Second, we found that both task demands and mental effort as measured by TEPR to be predictors of performance, specifically omission errors with no evidence of trying to complete the task (Table 3). We found that participants made more omission errors in PN case, followed by HF case when compared to UTI case, and had their TEPR dilated 110% more than participants that had no errors. These findings are in line with previous findings indicating that higher task demands lead to increase in mental effort that increases errors in various domains (Da Silva, 2014; Stetina, Groves, & Pafford, 2005) including physician–EHR interactions (Ariza, Kalra, & Potts, 2015; Mazur et al., 2016; Mosaly et al., 2016; Zhang & Walji, 2011).

Third, we found the increase in mental effort and task difficulty (click pattern) to be predictors of performance, specifically omission errors with evidence of trying but unable to complete the task (Table 3). We found that participants who made errors had significantly lower blink frequency (decreased by 50%) and exhibited frequent searching for orders (NID^m) click pattern behavior.

To further understand the task difficulty based on click behavior, we assessed the videos and found that participants utilized two different procedures, i.e., *templated* versus *non-templated*, while making orders within computerized physician order entry (CPOE). Templated procedure enabled

Table 3. Summary of the study results that had significant relationship with performance.

		Performance					
		No Errors		Omission Errors with *no* evidence of trying		Omission Error with evidence of trying	
Variable	Levels/Measures	*3-level*	*2-level*	*3-level*	*2-level*	*3-level*	*2-level*
Task Demand (3- or 2-level)* [count]	*UTI*	14	14	1	1	2	2
	PN	4	} 13	8^{+}	} 10	8	} 15
	HF	9		2		7	
Mental Effort [mean (SD)]	*Blink Freq. (#/minute)*	6.1(4)		5.3 (4)		3.2 (3) $^{+\ ++}$	
	TEPR (mm)	0.12 (0.13)		0.25 (0.14) $^{+\ ++}$		0.19 (0.13)	
Task Difficulty [count]	*Searching for orders (NID^m)*	11		12		$23^{+\ ++}$	

*3-levels: UTI, PN, HF; 2-levels: UTI, PN&HF

$^{+}$ 3-level task demand: p<0.05

$^{++}$ 2-level task demand: p<0.05

Table 4. Describes the %(*n*) of participants who followed *templated* vs. *non-templated* procedure and made (or not made) omission errors.

Errors	No Errors	Omission Errors with *no* evidence of trying	Omission Errors with evidence of trying
Followed *Templated* Procedure (63% [n=32])	50% (16/32)	28% (9/32)	22% (7/32)
Followed *Non-templated* Procedure (37%[n=19])	36% (7/19)	11% (2/19)	53% (10/19)

participants to access custom "Order Sets" via "Admissions" module (listing only the orders relevant to the selected reason [diagnosis] for admission). On the other hand, non-templated procedure enabled participants to access "Order Sets" directly, which required "free text search", for orders to be placed (thus, frequent search pattern). We found that participants who followed templated procedure had more omission errors with no evidence of trying to complete the task than participants who followed non-templated procedure; whereas participants who followed non-templated procedure had more omission errors with evidence of trying but unable to complete the task, than participants who followed templated procedure (Table 4).

Interestingly, these findings could be related back to the concepts of limited resource theory (Broadbent, 1958; Kahneman, 1973; Meister, 1976; Sperandio, 1971). That is, we believe that participants who used templated procedure experienced more mental effort (i.e., increased TEPR) that potentially contributed to omission errors as they "stopped" paying attention to our instructed list of tasks (attention tunneling or narrowing), and instead focused their attention on the listed orders presented in the Epic EHR system after they admitted the patient using "Admissions" module. For example, in PN case, 3 of the 17 participants using templated procedure checked "**high** risk for venous thromboembolism (VTE) prophylaxis" (high risk of blood clots) instead of "**low** risk for venous thromboembolism (VTE) prophylaxis" as instructed. These orders were presented in the EHR as a list format as "VTE Risk Category – High" being listed first followed by "Low Risk of VTE", with a check box in front of them. Participants chose the first box on the list by seeing the term VTE and not reading the complete phrase (term "High" was presented at the end of the phrase and term "Low" was presented at the beginning of the phrase for VTE risk), thus leading to an unconscious human error. Similar issues were also seen for orders like "sputum culture" (to find bacteria or fungi that are cause infection of the lungs or the airways leading to the lungs) for PN. Such challenges might be especially relevant during handoffs and cross-cover situations (Mazur et al., 2016; Mosaly et al., 2013). Our findings emphasize that some task-related characteristics that must be exercised by physicians when using templates in EHRs (e.g., decision rate per EHR display space, complexity of decisions, relationship between decisions) could enhance or constrain creative clinical thinking and promote automaticity (Hartzband & Groopman, 2008). In addition to task characteristics, our findings also highlight some technological characteristics (e.g., usability, functionality, search engines) to minimize mental effort as well as implementation of proper quality assurance (QA) check to assure self-regulation (or slow-down) of physicians' behavior to facilitate careful check of their work before approval.

On the other hand, participants who followed non-templated procedure were unable to complete instructed tasks despite the use of variety of keywords and multiple searches, which resulted in lower blink frequency due to increased mental effort via visual load. It is possible that these participants lacked proper training on keyword searches to effectively interact with EHR's database and/or sub-optimal functionality of EHR's search databases itself. For example, when using search keywords like "VTE", "high VTE, or "high risk venous", database would return only "**low** risk for VTE prophylaxis", resulting in a conscience human error (e.g., giving up search effort and selecting low instead of high VTE; which, itself seemed to some degree as a surprising behavior that compels further investigation). Similar behaviors were also seen for orders like "subcutaneous heparin" (used as anticoagulant injection to minimize blood clots), "Trans thoracic echocardiogram (TTE)" (a type of echocardiogram), and "IV Lasix" (to treat acute pulmonary edema), all exhibiting frequent multiple loops of repeated search behavior. Such challenges are often evident in real clinical settings where physicians, when unable to finds orders in EHRs, would delegate the task to their subordinates (i.e., to nurses or resident physicians) to complete the task. Zhang, Padman, and Levin (2014) also found similar issues with Epic EHR system while using templated versus non-templated procedures for order sets. They discussed advantages and disadvantages of one procedure over the other and recognize that usability improvement for the design of the order sets may not be the best solution in all applications. Our findings highlight the opportunity to use click patterns as a QA metric (e.g., flagging some interactions [great number of search patterns] as suboptimal with likelihood of error) in addition to usability improvement may improve performance. Similar efforts were done by Zheng, Padman, Johnson, and Diamond (2009) where they found that some of the navigational patterns in EHRs significantly deviated from the ideal patterns and therefore understanding these task difficulties or undesirable user behaviors characteristics may help in informing corrective actions such as focused user training or continued system reengineering.

Other highly reliable industries monitor and model human behaviors and interaction with computers to ensure safety (Loft et al., 2007; Parasuraman & Hancock, 2001; Sperandio, 1971). Thus, tracking users' behavior during EHR interaction could be useful to improve human performance (e.g., via constructive feedback) (Mazur et al., 2017), offer suggestions regarding suboptimal usability and functionality (Zhang et al., 2014) of EHRs and increase patient safety (Edwards, Moloney, Jacko, & Sainfort, 2008; Middleton et al., 2013). Overall, we found that most of our investigated relationships to be in line with previously established relationship within the HCI domain. However, the lack of relationship between task demands and task difficulty suggests that further methodological and empirical studies are needed to advance our knowledge on ways to quantify task difficulty during physicians' interactions with EHRs. Future research could also explore the ways to examine the relationship between mental effort and task difficulty, while considering potential tradeoffs between interaction speed and accuracy.

There are some limitations to this study, and thus caution should be exercised in generalizing our findings. First, study findings are based on 17 participants (resident physicians and medical students) and three clinical scenarios that are considered to be the most commonly performed tasks in the hospital environment (Pfuntner et al., 2013). However, all participants in the study were knowledgeable in performing these common tasks. Second, techniques employed to calculate TEPR, blink frequency and gaze velocity are challenging and are not perfect. For example, pupillary dilation is affected by change in illumination of the display, gaze etc. Any loss of the pupillary measures is coded as '0' in the raw data and hence the software has over or under estimated eye blinks, and change in participants' position (leaning toward and away from the display monitor), and repeatedly looking at the keyboard to perform typing may have affected the eye gaze, pupil and blink data. Therefore, we performed visual inspection of raw pupillary data to identify potential outliers and data that was considered invalid was discarded (and linearly interpolated for pupillary data), based on the recommendation provided by Beatty and Lucero-Wagoner (2000), Holmqvist et al. (2011) and Marshall (2005). Third, quantifications of task difficulty and performance are challenging and could be inexact. Therefore, two researchers independently reviewed the videos and coded mouse clicks (with discrepancies resolved during weekly meetings) before computerized sequences algorithms were used (Guo et al., 2016). We also had an experienced physician verifying our performance coding.

4.1. Conclusion

We found that physicians' performance during EHR interaction was negatively affected by task demands and increase in mental effort. The results also suggest that tracking task difficulty based on users' behavior (e.g., click patterns) could be used as a QA metric to predict performance. Overall, our findings are aligned with prior research indicating that high task demands evoke high mental effort and affect performance. However, the lack of relationship between task demand and task difficulty highlights a need for further methodological and empirical studies to advance our understanding of physicians' interactions with EHRs. In concert, our findings might bring the necessary knowledge, and urgency to help designers improve usability of EHRs, and healthcare organizations to choose appropriate EHRs and develop training methods to improve quality and patient safety.

Funding

This study was originally funded by the Department of Radiation Oncology, the University of North Carolina (UNC) and the UNC Health Care System. The data analysis of subjective and psychological measures of workload was partially supported by grants R18HS023458 and R21HS024062 from the Agency for Healthcare Research and Quality (AHRQ).The content is solely the responsibility of the authors and does not necessarily represent the official views of the AHRQ.

References

Ariza, F., Kalra, D., & Potts, H. W. (2015). How do clinical information systems affect the cognitive demands of general practitioners? Usability study with a focus on cognitive workload. *Journal of Innovation in Health Informatics*, *22*(4), 379–390. doi:10.14236/jhi.v22i4.85

Athènes, S., Averty, P., Puechmorel, S., Delahaye, D., & Collet, C. (2002). Complexity and controller workload: Trying to bridge the gap. *Proceedings of the 2002 International Conference on Human-Computer Interaction in Aeronautics* (pp. 56–60). Cambridge: Massachusetts Institute of Technology.

Ayaz, H., Willems, B., Bunce, B., Shewokis, P. A., Izzetoglu, K., Hah, S., … Onaral, B. (2010). Cognitive workload assessment of air traffic controllers using optical brain imaging sensors. Advances in understanding human performance. In *Neuroergonomics, human factors design, and special populations* (pp. 21–31), CRC Press Taylor & Francis Group.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. *Handbook of Psychophysiology*, *2*, 142–162.

Boehm-Davis, D. A., Casali, J. G., Kleiner, B. M., Lancaster, J. A., Saleem, J. J., & Wochinger, K. (2007). Pilot performance, strategy, and workload while executing approaches at steep angles and with lower landing minima. *Human Factors*, *49*(5), 759–772. doi:10.1518/001872007X230145

Bowman, S. (2013). Impact of Electronic Health Record Systems on Information Integrity: Quality and Safety Implications. *Perspectives in Health Information Management*, *10*(Fall), 1c.

Brennan, K., Kelly, D., & Arguello, J. (2014, August). The effect of cognitive abilities on information search for tasks of varying levels of complexity. In Proceedings of the 5th Information Interaction in Context Symposium (pp. 165–174). ACM.

Broadbent, D. E. (1958). The selective nature of learning. In *Perception and communication* (pp. 244–267). Elmsford, NY: Pergamon Press.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, *42*, 361–377. doi:10.1016/0301-0511(95)05167-8

Bruneau, D., Sasse, M. A., & McCarthy, J. D. (2002, April). The eyes never lie: The use of eye tracking data in HCI research. *Proceedings of the CHI 2002 Conference*, *2*, 25.

Byun, S. N., & Choi, S. N. (2002). An evaluation of the operator mental workload of advanced control facilities in Korea next generation reactor. *Journal of Korean Institute of Industrial Engineers*, *28*(2), 178–186.

Carswell, C. M., Clarke, D., & Seales, W. B. (2005). Assessing mental workload during laparoscopic surgery. *Surgical Innovation*, *12*(1), 80–90. doi:10.1177/155335060501200112

Chen, S., Epps, J., Ruiz, N., & Chen, F. (2011, February). Eye activity as a measure of human mental effort in HCI. *In Proceedings of the 16th international conference on Intelligent user interfaces* (pp.315–318). ACM.

Colle, H. A., & Reid, G. B. (2005). Estimating a mental workload redline in a simulated air-to-ground combat mission. *International Journal of Aviation Psychology*, *15*(4), 303–319. doi:10.1207/s15327108ijap1504_1

Cullen, R. H., Dan, C. S., Rogers, W. A., & Fisk, A. D. (2014). The effects of experience and strategy on visual attention allocation in an automated multiple-task environment. *International Journal of Human-Computer Interaction*, *30*(7), 533–546. doi:10.1080/10447318.2014.906158

Da Silva, F. P. (2014). Mental workload, task demand and driving performance: What relation? *Procedia-Social and Behavioral Sciences*, *162*, 310–319. doi:10.1016/j.sbspro.2014.12.212

De Rivecourt, M., Kuperus, M. N., Post, W. J., & Mulder, L. J. M. (2008). Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics*, *51*(9), 1295–1319. doi:10.1080/00140130802120267

De Waard, D. (1996). *The Measurement of Driver's Mental Workload* (Doctoral dissertation). Traffic Research Centre, University of Groningen, Haren, NL.

Edwards, P. J., Moloney, K. P., Jacko, J. A., & Sainfort, F. (2008). Evaluating usability of a commercial electronic health record: A case study. *International Journal of Human-Computer Studies*, *66*(10), 718–728. doi:10.1016/j.ijhcs.2008.06.002

Edwards, T., Gabets, C., Mercer, J., & Bienert, N. (2017). Task demand variation in air traffic control: implications for workload, fatigue, and performance. In: Stanton N., Landry S., Di Bucchianico G., Vallicelli A. (eds) *Advances in human aspects of transportation*. Advances in Intelligent Systems and Computing. (484) (pp. 91–102). Cham: Springer International Publishing.

Elliott, D. J., Young, R. S., Brice, J., Aguiar, R., & Kolm, P. (2014). Effect of hospitalist workload on the quality and efficiency of care. *JAMA Internal Medicine*, *174*(5), 786–793. doi:10.1001/jamainternmed.2014.300

Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology*, *56*(2), 171–184. doi:10.1016/j.ijpsycho.2004.11.003

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, *19*, 213–236. doi:10.2307/249689

Graber, M. L., Siegal, D., Riah, H., Johnston, D., & Kenyon, K. (2015). Electronic health record-related events in medical malpractice claims. *Journal of Patient Safety*, 1. doi:10.1097/PTS.0000000000000240

Guo, H., Gomez, S. R., Ziemkiewicz, C., & Laidlaw, D. H. (2016). A case study using visualization interaction logs and insight metrics to understand how analysts arrive at insights. *IEEE Transactions on Visualization and Computer Graphics*, *22*(1), 51–60. doi:10.1109/TVCG.2015.2467613

Hancock, P. A., & Szalma, J. L. (2003). Operator stress and display design. *Ergonomics in Design: the Quarterly of Human Factors Applications*, *11*(2), 13–18. doi:10.1177/106480460301100205

Hancock, P. A., Williams, G., & Manning, C. M. (1995). Influence of task demand characteristics on workload and performance. *The International Journal of Aviation Psychology*, *5*(1), 63–86. doi:10.1207/s15327108ijap0501_5

Hartzband, P., & Groopman, J. (2008). Off the record–avoiding the pitfalls of going electronic. *The New England Journal of Medicine*, *358*(16), 1656. doi:10.1056/NEJMp0802221

Hertzum, M., & Holmegaard, K. D. (2013). Perceived time as a measure of mental workload: Effects of time constraints and task success. *International Journal of Human-Computer Interaction*, *29*(1), 26–39. doi:10.1080/10447318.2012.676538

Hilburn, B. (2004). *Cognitive complexity in air traffic control: A literature review* (EUROCONTROL Experimental Centre Note No.04/04). Brussels, Belgium. Retrieved January 24, 2016, from http://www.chpr.nl/index_htm_files/CXLIT.pdf

Hill, R. G., Sears, L. M., & Melanson, S. W. (2013). 4000 clicks: A productivity analysis of electronic medical records in a community hospital ED. *The American Journal of Emergency Medicine*, *31*(11), 1591–1594. doi:10.1016/j.ajem.2013.06.028

Histon, J. M., & Hansman, R. J. (2002). *The impact of structure on cognitive complexity in air traffic control* (Rep. No. ICAT-2002-4). Cambridge, MA: MIT International Center for Air Transportation.

Hoffman, E., Pene, N., Rognin, L., & Zeghal, K. (2003, October). Introducing a new spacing instruction. Impact of spacing tolerance on flight crew activity. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *47*(1), 174–178. SAGE Publications. doi:10.1177/154193120304700136.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van De Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kantowitz, B. H. (1987). Mental workload. In P. A. Hancock (Ed.), *Human factors psychology* (pp. 81–121). Amsterdam, The Netherlands: North Holland Publishing Co.

Lee, K., Kerns, K., Bone, R., & Nickelson, M. (2001, December). Development and validation of the Controller Acceptance Rating Scale (CARS): Results of empirical research. *Proceedings of the 4th USA/Europe Air Traffic Management R&D Seminar*.

Liang, G. F., Lin, J. T., Hwang, S. L., Huang, F. H., Yenn, T. C., & Hsu, C. C. (2009). Evaluation and prediction of on-line maintenance workload in nuclear power plants. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *19*(1), 64–77.

Lin, H. S., & Stead, W. W.(Eds.). (2009). *Computational technology for effective health care: Immediate steps and strategic directions*. Washington, DC: National Academies Press.

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors*, *49*(3), 376–399. doi:10.1518/001872007X197017

Marshall, S. P. (2005). Assessing cognitive engagement and cognitive states from eye metrics. In D. Schmorrow (Ed.), *Foundations of augmented cognition* (pp. 312–320). Mahwah, NJ: Lawrence Erlbaum Associates.

May, J. G., Kennedy, R. S., Williams, M. C., Dunlap, W. P., & Brannan, J. R. (1990). Eye movement indices of mental workload. *Acta Psychologica*, *75*(1), 75–89. doi:10.1016/0001-6918(90)90067-P

Mazur, L. M., Mosaly, P. R., Moore, C., Comitz, E., Yu, F., Falchook, A. D., … Marks, L. B. (2016). Toward a better understanding of task demands, workload, and performance during physician-computer interactions. *Journal of the American Medical Informatics Association*, ocw016, *23*(6), 1113–1120.

Mazur, L. M., Mosaly, P. R., Tracton, G., Stiegler, M. P., Adams, R. D., Chera, B. S., & Marks, L. B. (2017). Improving radiation oncology providers' workload and performance: Can simulation-based training help?. Practical Radiation Oncology, *7*(5), e309–e316. doi:10.1016/j.prro.2017.02.005

Meister, D. (1976). *Behavioral foundations of systems development*. New York, NY: Wiley.

Merat, N., Antilla, V., & Luoma, J. (2005). Comparing the driving performance of average and older drivers: The effect of surrogate in-vehicle information system. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(2), 147–166. doi:10.1016/j.trf.2005.04.011

Middleton, B., Bloomrosen, M., Dente, M. A., Hashmat, B., Koppel, R., Overhage, J. M., … Zhang, J. (2013). Enhancing patient safety and quality of care by improving the usability of electronic health record systems: Recommendations from AMIA. *Journal of the American Medical Informatics Association*, *20(e1)*, e2–e8. doi:10.1136/amiajnl-2012-001458

Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica*, *27*, 84–92. doi:10.1016/0001-6918(67)90048-0

Moray, N., Dessouky, M. I., Kijowski, B. A., & Adapathya, R. (1991). Strategic behavior, workload, and performance in task scheduling. *Human Factors*, *33*(6), 607–629. doi:10.1177/001872089103300602

Mosaly, P. R., Mazur, L., & Marks, L. B. (2016, March). Usability evaluation of Electronic Health Record System (EHRs) using subjective and objective measures. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval* (pp. 313–316). ACM.

Mosaly, P. R., Mazur, L. M., Jones, E. L., Hoyle, L., Zagar, T., Chera, B. S., & Marks, L. B. (2013). Quantifying the impact of cross coverage on physician's workload and performance in radiation oncology. *Practical Radiation Oncology*, *3(4)*, e179–e186. doi:10.1016/j.prro.2013.02.007

Parasuraman, R., & Hancock, P. A. (2001). Adaptive control of mental workload. In B. H. Kantowitz (Eds.), *Stress, workload, and fatigue* (pp. 305–320). London, UK: Lawrence Erlbaum Associates, Inc.

Peute, L. W., & Jaspers, M. W. (2007). The significance of a usability evaluation of an emerging laboratory order entry system.

International Journal of Medical Informatics, *76*(2), 157–168. doi:10.1016/j.ijmedinf.2006.06.003

Pfuntner, A., Wier, L. M., & Stocks, C. (2013, September). *Most frequent conditions in U.S. hospitals, 2011* (HCUP statistical brief #162). Rockville, MD: Agency for Healthcare Research and Quality. Retrieved March 2, 2016, from http://www.hcup-us.ahrq.gov/reports/statbriefs/sb162.pdf

Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. *Encyclopedia of Human Computer Interaction*, *1*, 211–219.

Russ, A. L., Zillich, A. J., Melton, B. L., Russell, S. A., Chen, S., Spina, J. R., … Hawsey, J. M. (2014). Applying human factors principles to alert design increases efficiency and reduces prescribing errors in a scenario-based simulation. *Journal of the American Medical Informatics Association*, *21 (e2)*, e287–e296. doi:10.1136/amiajnl-2013-002045

Sparnon, E., & Marella, W. M. (2012). The role of the electronic health record in patient safety events. *PA Patient Saf Advis*, *9*(4), 113–121.

Sperandio, J. C. (1971). Variation of operator's strategies and regulating effects on workload. *Ergonomics*, *14*, 571–577. doi:10.1080/00140137108931277

Stetina, P., Groves, M., & Pafford, L. (2005). Managing medication errors: A qualitative study. *Medsurg Nurs*, *14*(3), 174–178.

Szalma, J. L. (2002, September). Individual differences in the stress and workload of sustained attention. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *46*(12), 1002–1006. Los Angeles, CA: SAGE Publications. doi:10.1177/154193120204601201.

Szalma, J. L., & Hancock, P. A. (2007). Task loading and stress in human–computer interaction: Theoretical frameworks and mitigation strategies. In *Handbook for human–computer interaction in interactive systems, 2nd Edition* (pp. 115–132). New York: Erlbaum Taylor & Francis Group.

Vicente, K. J. (1999). Cognitive work analysis: Toward safe, productive, and healthy computer-based work. Mahwah, NJ: Lawrence Erlbaum Associates.

Young, G., Zavelina, L., & Hooper, V. (2008). Assessment of workload using NASA Task Load Index in perianesthesia nursing. *Journal of PeriAnesthesia Nursing*, *23*(2), 102–110. doi:10.1016/j.jopan.2008.01.008

Yurko, Y. Y., Scerbo, M. W., Prabhu, A. S., Acker, C. E., & Stefanidis, D. (2010). Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *Simulation in Healthcare*, *5*(5), 267–271. doi:10.1097/SIH.0b013e3181e3f329

Zhang, J., & Walji, M. F. (2011). TURF: Toward a unified framework of EHR usability. *Journal of Biomedical Informatics*, *44*(6), 1056–1067. doi:10.1016/j.jbi.2011.08.005

Zhang, P., & Galletta, D. F. (2006). *Human-computer interaction and management information systems: Foundations*. Armonk, NY: ME Sharpe.

Zhang, Y., Padman, R., & Levin, J. E. (2012). Reducing provider cognitive workload in CPOE use: Optimizing order sets. *Studies in Health Technology and Informatics*, *192*, 734–738.

Zhang, Y., Padman, R., & Levin, J. E. (2014). Paving the COWpath: Data-driven design of pediatric order sets. *Journal of the American Medical Informatics Association*, *21(e2)*, e304–e311. doi:10.1136/amiajnl-2013-002316

Zheng, B., Jiang, X., Tien, G., Meneghetti, A., Panton, O. N. M., & Atkins, M. S. (2012). Workload assessment of surgeons: Correlation between NASA TLX and blinks. *Surgical Endoscopy*, *26*(10), 2746–2750. doi:10.1007/s00464-012-2268-6

Zheng, K., Padman, R., Johnson, M. P., & Diamond, H. S. (2009). An interface-driven analysis of user interactions with an electronic health records system. *Journal of the American Medical Informatics Association*, *16*(2), 228–237. doi:10.1197/jamia.M2852

About the Authors

Prithima Mosaly (Prithima_mosaly@med.unc.edu) is a human factors and ergonomics specialist with an interest in utilization of pupillary and eye behavior in assessing cognitive workload/performance during human–computer interaction (HCI) in healthcare; she is a Research Assistant Professor in the Department of Radiation Oncology at the University of North Carolina at Chapel Hill.

Lukasz Mazur (lukasz_mazur@med.unc.edu) is a healthcare quality improvement specialist with an interest in healthcare engineering management, simulation-based training, hospital design with focus on Lean-Evidence-Based-Design integration; and cognitive workload during HCI; he is an Assistant Professor and Director of Division of Healthcare Engineering (DHE) in the Department of Radiation Oncology at the University of North Carolina at Chapel Hill.

Fei Yu (feifei@live.unc.edu) is a health informatics specialist and interested in usability testing of electronic health record systems, data analytics and visualization; she is a Research Associate in the Laboratory of Applied Informatics Research at the University of North Carolina at Chapel Hill.

Hua Guo (hua_guo@brown.edu) is a computer scientist with an interest in visualization and human computer interaction and in particular in informing and automating visualization design and evaluation using human-centered approaches; she is a 5th year PhD student in the Department of Computer Science at Brown University.

Derek Merck (derek_merck@brown.edu) is an expert in diagnostic Imaging; he is an Associate Professor and the Director of Computer vision and image analysis in the department of radiation oncology at Brown University.

David Laidlaw (david_laidlaw@brown.edu) is a computer scientist and his research centers on applications of visualization, modeling, computer graphics, and computer science to other scientific disciplines; he is a Professor of the Department of Computer Science at Brown University.

Lawrence Marks (marks@med.unc.edu) is a radiation oncologist with clinical practice focusing on cancers of the breast and lung; he is passionate about human factors, quality improvement, Lean and patient safety; he is a distinguished Professor and Chairman of the Department of Radiation Oncology at the University of North Carolina at Chapel Hill.

Javed Mostafa (jm@unc.edu) is a an expert in multimedia information retrieval, personalization and user modeling, cyber infrastructure for research and learning, and health informatics; he is a Professor and Director of School of Information and Library Science at the University of North Carolina at Chapel Hill.