Task Switching and Accuracy of Rapid Electrocardiogram screening for ST-Elevation Myocardial Infarction by Emergency Medicine Providers

A thesis

Submitted by

Dr. William Edward Soares III, MD.

In partial fulfillment of the requirements
For the degree of

Master of Science

In

Tufts Clinical and Translational Science Institute (CTSI)

TUFTS UNIVERSITY

Sackler School of Graduate Biomedical Sciences

August 2015

Advisor

Thesis Chair: Dr. Rebecca Blanchard, PhD.
Project Mentor: Dr. Timothy Mader, MD.
Statistical Mentor: Lori Lyn Price, MS.

Abstract

Interruptions are frequent in healthcare and are thought to contribute to medical errors. However, pausing workflow to address clinically significant interruptions, known as task switching, is an important aspect of emergency care. While prior studies have focused on primary task errors, no study has evaluated the association of task switching on accuracy of interpretation of critical interruptions.

A quasi-experimental trial utilizing task switching and uninterrupted simulations was created. Participants first completed a task-switching module, viewing patient presentations with clinical interruptions requiring interpretation every minute. Participants then completed an uninterrupted module, in which patient presentations and clinical stimuli were completed sequentially and without interruption. The primary outcome was the percent of electrocardiograms (ECGs) correctly interpreted for ST elevation myocardial infarction (STEMI) during the task-switching module compared to uninterrupted module.

Overall, 35 emergency providers completed the study. There was no significant difference in mean percent correct (accuracy) of ECG interpretation for STEMI in the task-switching scenario compared with the uninterrupted scenario (Mean 0.89, 0.91 respectively, Paired T test $p=0.21$). Logistic regression using generalized estimating equations revealed attending physician status (OR 2.56, CI 1.66-3.94, $p<0.01$) and inferior STEMI (OR 0.08, CI 0.04-0.14, $p<0.01$) were associated with increased and decreased odds of correct ECG interpretation, respectively. Further, self reported confidence was associated with increased odds

of correct ECG interpretation in the task-switching module, but not in the uninterrupted module. (Interaction p=0.02)

Results suggest that not all interruptions lead to error; factors such as provider experience, type of interruption and confidence in interpretation all influence the probability of error. Further research, both to identify factors associated with medical errors interpreting interruptions as well as strategies to avoid errors is needed.

Acknowledgements

I would like to thank my thesis team, Dr. Rebecca Blanchard, Dr. Timothy Mader and Lori Lyn Price for their support and guidance throughout the thesis project. I would also like to thank the Tufts CTSI faculty, especially Dr. David Kent and Dr. Jess Paulus for their mentorship during the masters program as well as their critical analysis of the project. Additionally, I would like to thank the staff associated with WaveMaven ECG website for allowing us to utilize their electrocardiograms for our project. Finally, I want to thank the residents and attending physicians of Baystate Medical Center, whose participation was critical to the success of the project.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BMC – Baystate Medical Center
CXR – Chest Radiograph
ECG – Electrocardiogram
GEE – Generalized Estimating Equations
STEMI – ST elevation myocardial infarction

Introduction

Interruptions, events that transiently break attention from a primary task, are common in emergency medicine; on average, emergency providers are interrupted between 5 and 30 times per hour (1-4). The Institute of Medicine's landmark 2000 paper, *To Err is Human,* argued interruptions play a role in medical errors. (5) According to cognitive theory, interruptions force working cognitive resources temporarily away from a primary task, leading to capacity and structural interference. This interference produces information overload, disrupting normal cognitive processing and causing errors. (6) Studies in psychology (7) aviation (8) and tactical decision-making (9) have corroborated the association of interruptions and errors in the primary task.

However, the relationship of healthcare interruptions and clinical errors remains uncertain. (10) While observational studies involving medication order entry (11) and pharmacy dispensing (12) suggest associations between interruptions and medical error, simulated experimental trials evaluating interruptions during medication order entry (13), clinical decision-making (14) and surgical procedures (15) have failed to demonstrate a relationship to primary task errors.

Further, most studies do not account for the importance of interruptions in the clinical setting. Patient monitor alarms, urgent phone calls, and interpretation of critical results are all examples of clinical interruptions that provide important information that must be quickly and accurately assessed. We found no studies that evaluate the relationship of task switching (interrupted workflow involving switching between a primary task and secondary interrupting tasks) and errors made in interpretation of clinically important interruptions.

The goal of the study is to explore the relationship of interrupted, task-switching workflow on interpretation of clinically important interruptions. Specifically, the current study investigates emergency providers accuracy interpreting electrocardiograms (ECG's) for ST-Elevation myocardial infarction (STEMI), a common critical action performed by emergency providers, in a simulated interrupted environment.

Methods

A quasi-experimental crossover design, was created. The trial incorporated a simulated task-switching environment as well as an uninterrupted environment to assess emergency providers accuracy of ECG interpretation for STEMI. The study was approved by the Baystate Medical Center IRB.

Study Population

The study population included intern, resident and attending physicians from the 3-year emergency medicine residency program at Baystate Medical Center (BMC) BMC is a tertiary care hospital and STEMI receiving center located in Springfield MA. Participants were recruited from November 2014 until January 2015. There were no exclusion criteria.

Study Design:

The study was comprised of two sections; a task-switching module designed to simulate interrupted workflow in the emergency department and an uninterrupted module designed to mirror a test or examination setting. Participants completed both the task switching and uninterrupted modules during a scheduled 1-hour session. In an attempt to maximize external validity and avoid priming bias (memory of a practiced stimulus influencing subsequent stimuli), all participants completed the task-switching module first and the uninterrupted module second. Verbal consent was obtained prior to the start of the simulation; participants were told the goal of the study was to evaluate how interruptions affect clinical decision-making.

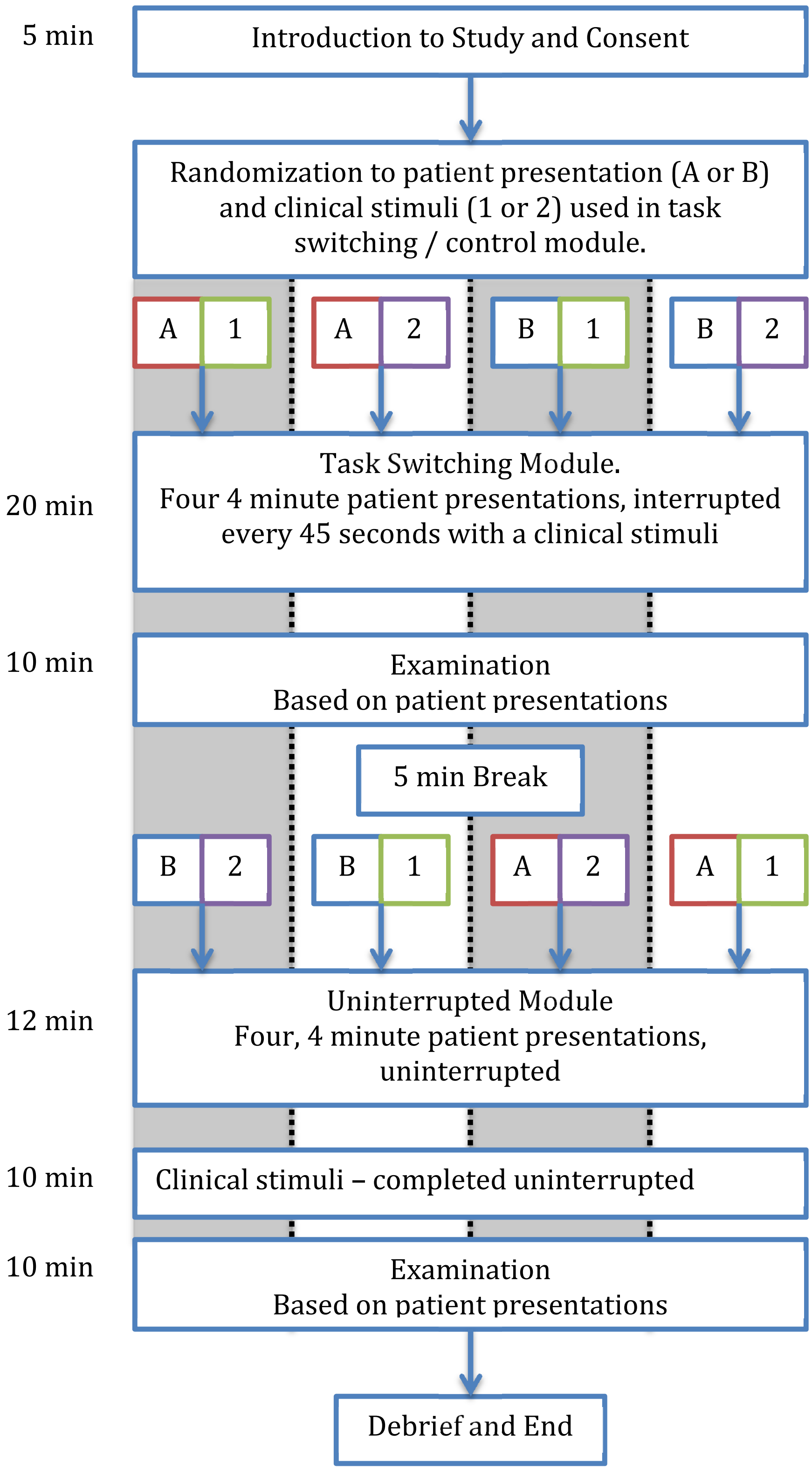
In the task-switching module, participants viewed four pre-recorded patient presentations, with the stated goal of remembering as many details from each presentation as possible. During the presentations, participants were interrupted every minute with various clinical stimuli, including chest x-rays

(CXR), "triage" ECG's and critical laboratory values from unrelated patients. Participants were asked to quickly evaluate the interruption for any actionable findings and then return to the patient presentations. To avoid revealing the primary outcome, actionable findings were openly defined as a discovery on the clinical stimuli that would require the patient be immediately brought from triage for clinical evaluation. During interruptions, patient presentations paused for 15 seconds to allow interpretation of stimuli. If needed, participants were allowed more time to interpret the stimuli, however, the patient presentations resumed automatically after 15 seconds. Once interpreted, participants placed clinical stimuli face down to the side.

In the uninterrupted module, participants listened to an additional four pre-recorded patient presentations sequentially and without interruption. Participants then evaluated a second group of clinical stimuli (CXR, ECG, laboratory values) for critical findings, however, without formal time constraint so as to mimic an examination setting.

After finishing each module, participants completed a written examination, evaluating their memory of patient presentation details as well as their ability to formulate a differential diagnosis on each patient. Figure 1 illustrates participant's flow through the simulation.

Figure 1: Diagram of participant flow through simulation.



Presentation Videos

Two presentation videos (A and B) were developed for the study by the principal investigator, WS, and evaluated for content validity by author and content expert TM. Each video consisted of four, 4-minute presentations of fictitious emergency medicine patients by a simulated provider. Patient presentations were designed to be complex with multiple possible diagnoses. To ensure equal difficulty, presentations were matched by number of items in the history of present illness, past medical history, past surgeries, medications, allergies and physical exam. Table 1 highlights factors matched in patient presentations. Videos were recorded using Camtasia for Mac software package.

Table 1. Factors Matched for Patient presentation

	Group A	Group B
Type	Gen Adult	Gen Adult
Problem	Chest Pain	Abd pain
History	4	4
Surgeries	1	1
Meds	6	5
Allergies	0	0
Exam	8	6
	Group A	Group B
Type	Elderly	Elderly
Problem	AMS	Back Pain
History	5	5
Surgeries	2	2
Meds	5	5
Allergies	1	1
Exam	8	8
	Group A	Group B
Type	Pediatrics	Pediatrics
Problem	Abd. Pain	Vomiting
History	1	1
Surgeries	1	1
Meds	1	1
Allergies	1	1
Exam	5	6
	Group A	Group B
Type	Trauma	Trauma
Problem	MVC	Assault
History	3	3
Surgeries	0	1
Meds	2	2
Allergies	1	1
Exam	6	6

Examinations testing memory of patient presentation details as well as ability to synthesize a differential diagnosis were created. Each examination consisted of 20 questions. Four questions per patient involved recall of specific details including chief complaint, past medical history, medications and physical exam abnormalities. The last question asked participants to create a differential diagnosis for the case. Each test was graded out of a total of 36 points.

Clinical Stimuli Packets

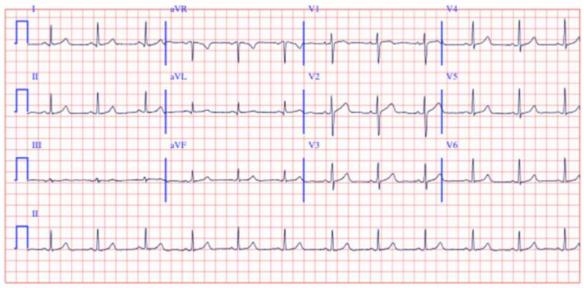
Two unique clinical stimuli packets (1 and 2) were developed for the study by the primary investigator, WS. Each clinical stimuli packet contained 13 ECG's, 4 CXR's, and 3 fictitious laboratory values.

Permission was obtained from the WaveMaven ECG teaching website to utilize their published ECG's in the current study. WaveMaven is a database of thousands of de-identified online ECGs validated by cardiology diagnosis and rated by difficulty. *(Nathanson L A, McClennen S, Safran C, Goldberger AL. ECG Wave-Maven: Self-Assessment Program for Students and Clinicians. http://ecg.bidmc.harvard.edu.)*

Each stimuli packet contained unique ECG's, matched by diagnosis and difficulty rating. Specifically, each packet contained 5 ECG's with STEMI, 4 normal ECG's and 4 ECG's with non-critical findings (e.g. benign early repolarization, left ventricular hypertrophy, right bundle branch block). De-identified CXR's and laboratory values, both normal and abnormal, were added to conceal the primary outcome of interest. Finally, each clinical stimulus included a section for participants to rate their confidence in their answer on a Likert scale from (1) "None" to (5) "Sure". Figure 2 shows examples of clinical stimuli.

Figure 2. Representative examples from Clinical Stimuli Packet A of an ECG stimulus, a CXR stimulus, and a Lab value stimulus.

Actionable Finding? YES / NO
Finding: __________
Confidence in Answer:
1 2 3 4 5
none sure



Actionable Finding? YES / NO
Finding: __________
Confidence in Answer:
1 2 3 4 5
none sure

http://radiopaedia.org

Actionable Finding? YES / NO
Finding: __________
Confidence in Answer:
1 2 3 4 5
none sure

Na	140
K	4.5
Cl	104
C02	25
BUN	10
Cr	0.8
Glu	126

Validation of Clinical Stimuli

Once created, stimuli packets were piloted in a cohort of 12 attending and senior resident emergency physicians to ensure concordance of difficulty between tests. These providers were not affiliated with Baystate Medical Center and did not participate in the study. They were asked to evaluate selected ECG's for STEMI and rate their confidence for each ECG interpretation (5 point Likert range 0%-100%). Table 2 demonstrates mean scores of participants for ECG's in test A and b with associated mean confidence levels. No difference in mean percent correct was found overall (Test 1 mean= 0.91, SD=0.1, Test 2 median = 0.89, SD = 0.08 Paired T Test p=0.39).

Table 2. Mean participant scores and confidence in validation group.

	Test A	Test B	
STEMI *mean(SD)*	0.87 (0.13)	0.82(0.13)	
Confidence *mean*	0.86	0.8	
non STEMI *mean(SD)*	0.97(0.08)	0.96(0.06)	
Confidence *mean*	0.83	0.82	
Total *mean(SD)*	0.91(0.1)	0.89(0.08)	
Confidence *mean*	0.84	0.81	

Given that participants were senior residents and attending physicians evaluating specifically for STEMI in an untimed environment and still only scored 80%-90%, it was thought the test would be adequately difficult for the proposed study.

Randomization

To control for potential unmeasured differences in difficulty of patient presentations and clinical stimuli, participant were randomized to which patient presentation video (A or B) and which clinical stimuli (1 or 2) they received in the task-switching module (the other patient presentation video and clinical stimuli were administered in the uninterrupted module.) Block randomization into groups of 4 was performed using online randomization software, http://www.randomizer.org/. Participants were not randomized by position (Table 3).

Table 3. Randomization of participants overall and by position.

Overall			
	Scenario A	Scenario B	
Stimuli 1	8	9	17
Stimuli 2	9	9	18
	17	18	35
Interns			
	Scenario A	Scenario B	
Stimuli 1	3	0	3
Stimuli 2	2	3	5
	5	3	8
Senior Res			
	Scenario A	Scenario B	
Stimuli 1	2	5	7
Stimuli 2	4	1	5
	6	6	12
Attending			
	Scenario A	Scenario B	
Stimuli 1	3	4	7
Stimuli 2	3	5	8
	6	9	15

Primary Study Outcome

The primary study outcome was defined as difference in accuracy of ECG interpretation for STEMI in the task-switching module compared to the uninterrupted module. Accuracy was defined as the mean sum of true positive and true negative interpretations divided by the total number of ECG's, or mean percent correct. A true positive interpretation was coded if participants correctly indicated a critical finding on the STEMI ECG with a corresponding written diagnosis of STEMI (or equivalent). A true negative interpretation was coded if participants correctly indicated no critical findings on normal / non-critical ECG's, or if participants indicated a critical finding unrelated to the primary outcome (such as atrial fibrillation or right bundle branch block).

Secondary Outcomes

Secondary study outcomes included differences in mean sensitivity and mean specificity of ECG interpretation for STEMI. Sensitivity and specificity were calculated for each participant during each module; means across participants were then used in final analysis. Additionally, recall of patient presentations, measured as the mean difference in written patient scenario examination scores (scenario exam), between task-switching and uninterrupted modules was evaluated.

Explanatory variables with hypothesized relationships to accuracy of ECG interpretation of STEMI were defined *a priori* for evaluation in a generalized logistic model. These covariates included provider experience (intern, senior resident, attending), score on written patient evaluations, type of ECG (non-STEMI, anterior / lateral STEMI, inferior/posterior STEMI) and confidence in stimuli interpretation (Likert, 1, none to 5, sure).

Statistical Analysis

Data were initially summarized using nonparametric descriptive statistics and hypothesis testing. However, on further analysis, results of parametric testing were similar to nonparametric counterparts and offered a more easily interpretable format. Therefore, data were presented using parametric descriptive statistics and hypothesis testing.

Paired sample hypothesis testing using Paired T tests was performed to evaluate differences in overall ECG accuracy, as well as differences in mean sensitivity, mean specificity and memory of patient presentations.

Given limited sample size and multiple predefined covariates, repeated measures logistic regression using odds of correct interpretation of each individual ECG as a binary outcome was utilized. Generalized estimating equations (GEE) with the exchangeable working correlation matrix structure were employed to account for non-independence of outcomes in the logistic regression model. Clinically relevant explanatory variables defined *a priori* were incorporated into the GEE model. Confidence demonstrated nonlinearity and was transformed for analysis into a dichotomous variable (low confidence = Likert 1-3, high confidence = Likert 4-5).

To evaluate for the presence of effect modification, variables including position, type of ECG and self reported confidence in interpretation were selected *a priori* for analysis as interaction terms with interrupted status (task switching module versus uninterrupted module) in the GEE model.

All statistical analyses were performed using SAS software 9.4 (SAS Institute, Inc., Cary, NC)) and R statistical software. 2014. (R Foundation for Statistical Computing, Vienna, Austria). P-values < 0.05 were considered statistically significant.

<u>GEE Model Building Process</u>

Initially, a generalized logistic regression model was created to further explore cofactors associated with ECG interpretation. As such, individual ECG interpretation for STEMI was used as the dichotomous outcome, grouped by participant (total n=910 with 26 ECG's for each of the 35 participants). Because the values of the outcome are not statistically independent, Generalized Estimating Equations were utilized to account for correlated clusters of participant data using a compound symmetry correlation matrix.

Covariates included interruption group (dichotomous, task switching or uninterrupted), position (categorical, intern, senior resident, attending physician), type of ECG (categorical, normal or no STEMI, Anterior STEMI, Inferior STEMI), scenario exam scores (continuous score) and Confidence of ECG interpretation (continuous score). Table 4 below demonstrates the counts of correct and incorrect ECG interpretation, stratified by covariates, in the long data set form.

Table 4: Summative data of correct / incorrect interpretation of individual ECG's, total and stratified by covariates. Scenario Exam is reported as mean, standard deviation. Confidence is reported both as mean, standard deviation as well as count. All other data is reported as count.

Variable	Correct ECG (n=817)	Incorrect ECG (n=93)
Interruption		
Task Switching	404	51
Uninterrputed	413	42
Position		
Intern	177	31
Senior Res	275	37
Attending	365	25
Type of ECG		
Normal	532	28
Anterior STEMI	201	9
Inferior STEMI	84	56
Scenario Exam (SD)	26.6(4.1)	26.5(4.2)
Confidence (SD)	4.06 (0.91)	3.41 (0.96)

Using restricted cubic spline method, we evaluated the two continuous covariates, Scenario Exam Scores and Confidence, for linearity. Scenario Exam Scores demonstrated linearity while Confidence appeared non linear.

Further, the two continuous covariates, Scenario Exam Scores (below as patient scenario scores) and Confidence, were evaluated for normality with QQ plots .

The QQ plot for Scenario Exam scores suggests a slight left skew. We attempted transformation of the Scenario Exam Score by squaring results (x^2). This did not appear to significantly improve the distribution according to the QQ plot. Further, Scenario Exam Score and the transformed Scenario Exam Score demonstrated linearity using restricted cubic splines method. Therefore, for simplicity, Scenario Exam scores were kept as a continuous variable.

Regarding Confidence, the QQ plot demonstrates a heavy tail left skew distribution, indicating many points at the extreme rather than the center. We attempted to correct by transformation (x^2 and x^3), however, this did not appear to improve distribution. Further, as stated above, confidence demonstrated non-linearity using restricted cubic splines method. Therefore, it was decided to treat confidence as a non-continuous variable. Given the relatively limited number of outcomes (90), ordinal transformation did not seem appropriate Therefore, we chose to dichotomize confidence into two scores, low confidence (1-3) and high confidence (4-5) based on the distribution of scores by figure 3. Table 5 summarizes the new Confidence dichotomous variable compared to prior continuous and ordinal grouping.

Figure 3. Interaction of confidence by interruption of probability of correct ECG interpretation

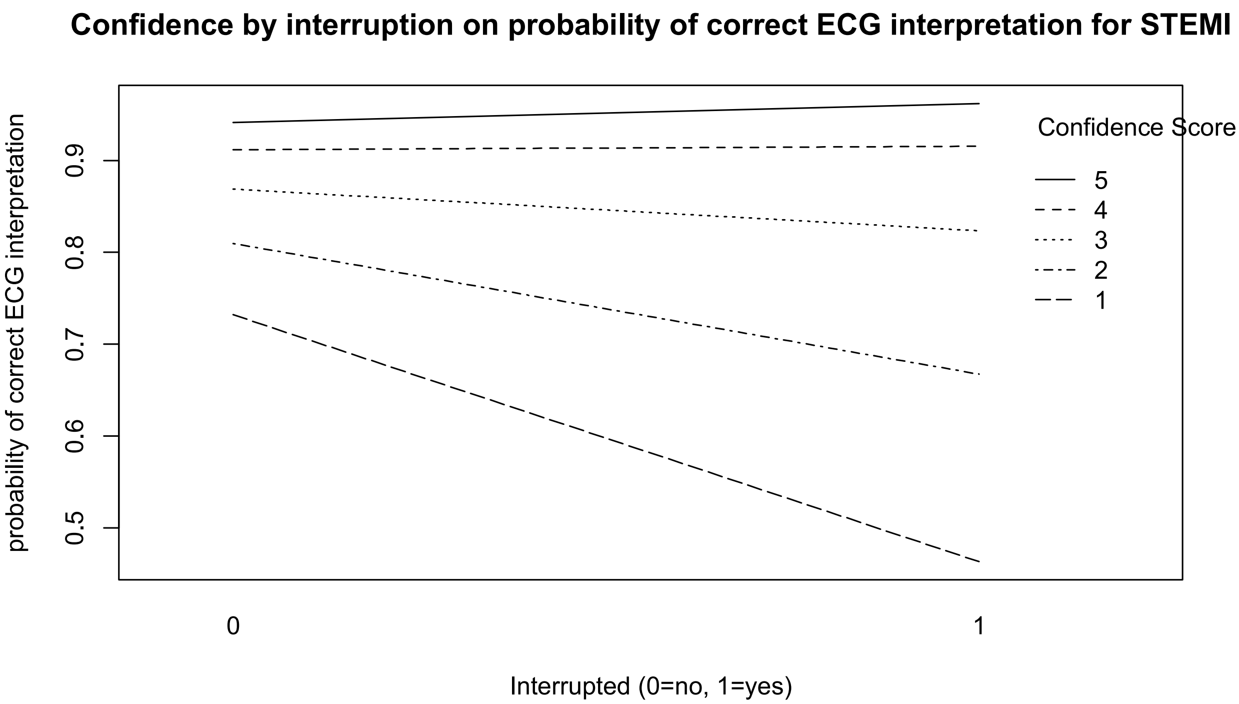


Table 5. Confidence variable as continuous, ordinal and dichotomous

Variable	Correct ECG (n=817)	Incorrect ECG (n=93)
Confidence (SD)	4.06 (0.91)	3.41 (0.96)
Confidence (ordinal)		
Low 1	8	4
2	34	9
3	168	35
4	297	35
High 5	310	10
Confidence (Dichot)		
Low (1-3)	210	48
High (4-5)	607	45

The final full GEE logistic regression model is as follows:

Logit(Correct Score)=Interrupted (0,1)+Position (0,1,2)+Type of ECG (0,1,2)+ Scenario Exam (continuous) +Confidence of ECG (0,1)

Power Analysis

Using Paired T-test, with a two-tailed alpha of 0.05, power analysis estimated that 33 participants would allow a power of 0.9 to demonstrate a 0.1 difference in accuracy, or, an approximate difference of 2 ECG's in 26 presented for interpretation, with a standard deviation of 0.20.

Results

Thirty-five emergency providers completed the study, including 8 intern, 12 senior resident and 15 attending physicians. Years in emergency medicine ranged from less than 1 year to 47 years (median 3 years, 25^{th},75^{th} percentile= 2,10).

Performance of Study Materials

Irrespective of module (task switching or uninterrupted), participants mean accuracy on ECG's in clinical stimuli 1 was not significantly different than mean accuracy scores on ECG's in stimuli 2, (mean 0.9, SD 0.08, mean 0.9, SD 0.08 respectively, p = 0.91). Mean scores on the scenario exams also did not demonstrate significant differences across modules (Scenario exam A = 0.73, SD 0.12, Scenario Exam B = 0.75,, SD 0.11 Paired T Test p = 0.23).

Primary Outcome

Emergency providers mean accuracy on identification of STEMI by ECG was not significantly different during task switching than during uninterrupted modules (Task Switching=0.89, SD = 0.08, Uninterrupted=0.91, SD = 0.08, p=0.21) Table 6 demonstrates overall mean scores as well as scores stratified by clinical position.

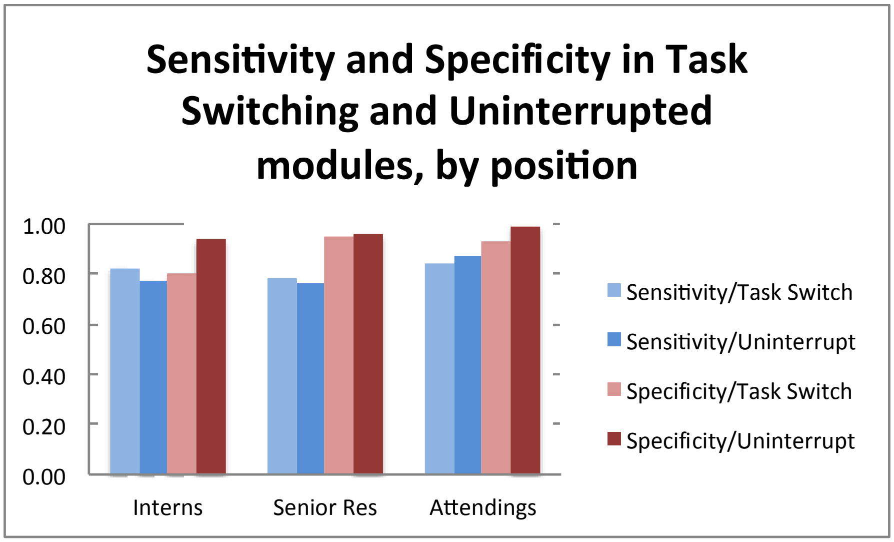
Table 6. Mean accuracy in task switching and uninterrupted modules, overall and stratified by position. Hypothesis testing using Paired T test.

	Task Switching	Uninterrupted	p-value
All Participants (n=35)			
Mean Accuracy	0.89	0.91	0.21
SD	0.08	0.08	
Interns (n=8)			
Mean Accuracy	0.82	0.88	0.17
SD	0.08	0.09	
Senior Resident (n=12)			
Mean Accuracy	0.88	0.88	0.81
SD	0.09	0.08	
Attendings (n=15)			
Mean Accuracy	0.93	0.94	0.38
SD	0.05	0.05	

Secondary Outcomes

Overall, mean sensitivity of ECG interpretation for STEMI in the task-switching module was 0.82 (SD=0.13) compared with mean sensitivity in the uninterrupted module of 0.81 (SD = 0.18, p=0.84). Specificity of ECG interpretation for STEMI trended toward decreasing in the interrupted scenario (0.90, SD = 0.19) compared to the uninterrupted scenario (0.97, SD=0.07, p=0.07). Stratified by clinical position, there were no significant differences in mean sensitivity between task-switching and uninterrupted scenarios. Regarding mean specificity, interns demonstrated a trend toward increased specificity in the uninterrupted clinical environment, (task switching = 0.8, SD = 0.13, uninterrupted=0.94, SD = 0.09, p=0.08) however, this was not seen in senior residents and attending physicians (Figure 4)

Figure 4: sensitivity and specificity of participants, by position in the interrupted and uninterrupted clinical environments.



Mean scores on scenario exams were not significantly different in the task switching and uninterrupted modules (task switching = 26.7 out of 36, SD = 4.24, uninterrupted=26.5 out of 36, SD 4.18, p=0.80). (Table 7)

Table 7. Mean scores on scenario exams, overall and by position

	Task Switching	Uninterrupted	p-value
All Participants (n=35)			
Mean Score (out of 36)	26.7	26.5	0.80
SD	4.2	4.1	
Interns (n=8)			
Mean Score (out of 36)	26.8	26.3	0.82
SD	4.7	2.1	
Senior Resident (n=12)			
Mean Score (out of 36)	26.0	28.0	0.19
SD	4.1	5.2	
Attendings (n=15)			
Mean Score (out of 36)	27.1	25.3	0.08
SD	4.3	3.8	

Regression Analysis

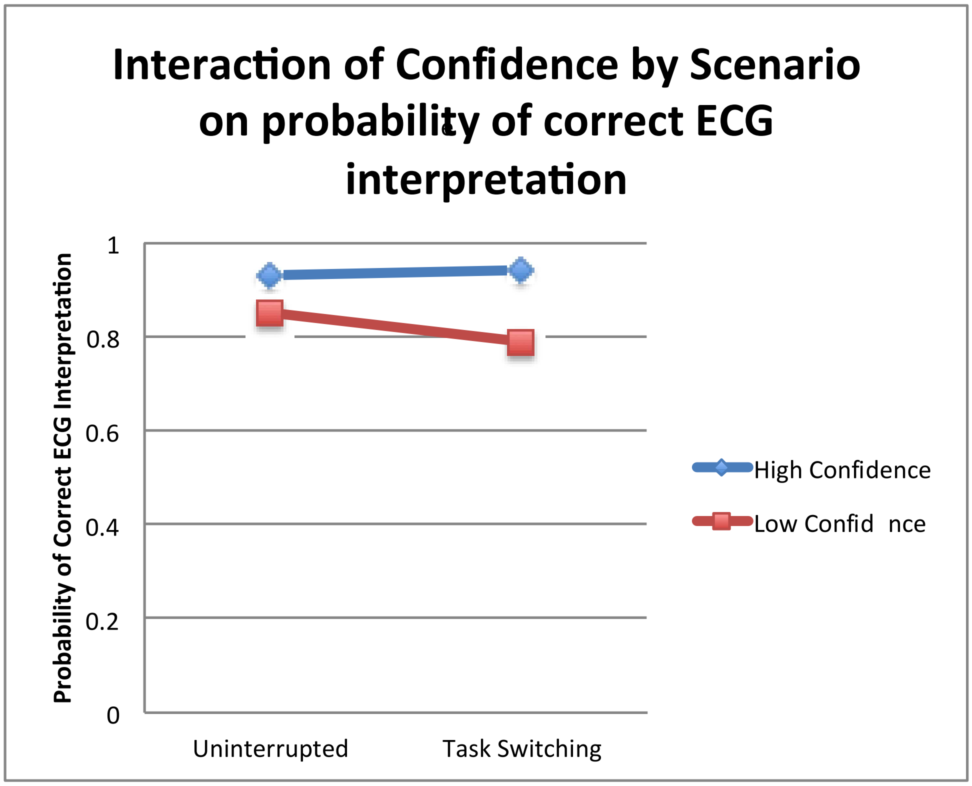
GEE Logistic regression analysis found that an interrupted environment was not significantly associated with the odds of correctly interpreting ECG's for STEMI (OR 0.81, CI 0.58-1.12, p=0.32). Covariates that were related to correct interpretation of ECG included attending clinical position (OR 2.56, CI 1.66-3.94, p<0.01) and self reported confidence in individual ECG interpretation (OR 3.10, CI 2.14-4.50, p<0.01). Presence of an inferior STEMI was associated with decreased odds of correct ECG interpretation (OR 0.08, CI 0.04-0.14, p<0.01). Table 8 demonstrates univariate and full model GEE logistic regression results.

Table 8: Logistic Regression (univariate and full model) using generalized estimating equations for cofactors associated with correct ECG interpretation for STEMI. Low confidence was defined as a self reported Likert score of 1-3, whereas high confidence was defined as a self reported Likert score of 4-5.

	GEE Univariate			GEE Full Model		
Variable	OR	CI	p value	OR	CI	p value
Scenario						
Uninterrupted (base)	1.00			1.00		
Interrupted	0.81	0.58-1.12	0.32	0.80	0.51-1.24	0.31
Position						
Intern (base)	1.00			1.00		
Senior Res	1.30	0.80-2.13	0.26	1.29	0.68-2.47	0.44
Attending	2.56	1.66-3.94	<0.01	2.40	1.42-4.05	<0.01
Type of ECG						
Normal (base)	1.00			1.00		
Anterior STEMI	1.17	0.44-3.13	0.67	0.78	0.30-2.03	0.61
Inferior STEMI	0.08	0.04-0.14	<0.01	0.06	0.03-0.11	<0.01
Mean Scenario Exam	1.01	0.96-1.05	0.83	1.01	0.96-1.06	0.62
Confidence						
Low (1-3) (base)	1.00			1.00		
High(4-5)	3.10	2.14-4.50	<0.01	3.68	2.26-6.01	<0.01

Regarding interaction effects between interrupted environment (task-switching or uninterrupted modules) and covariates, the interaction of interrupted environment with confidence reached significance (p=0.02). In an interrupted environment, decreased confidence was associated with lower odds of correct ECG interpretation, compared with an uninterrupted environment. Position and ECG type demonstrated no significant interaction in predicting correct ECG interpretation. Figure 5 demonstrates the interaction of confidence and interrupted environment on the estimated probability of correct ECG interpretation.

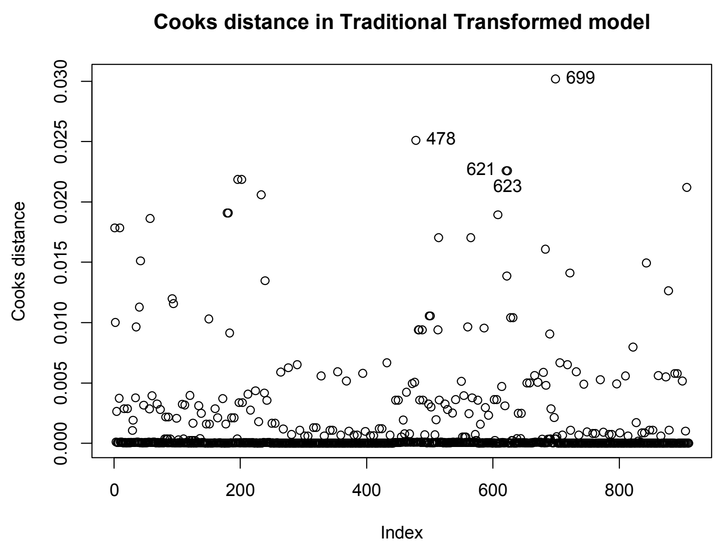
Figure 5: Interaction effect of confidence and interrupted environment on the estimated probability of correct ECG interpretation for STEMI. Low confidence (red) refers to Likert scores of 1-3 whereas high confidence (blue) refers to Likert scores of 4-5. The p value for the interaction was 0.02



Residual Analysis

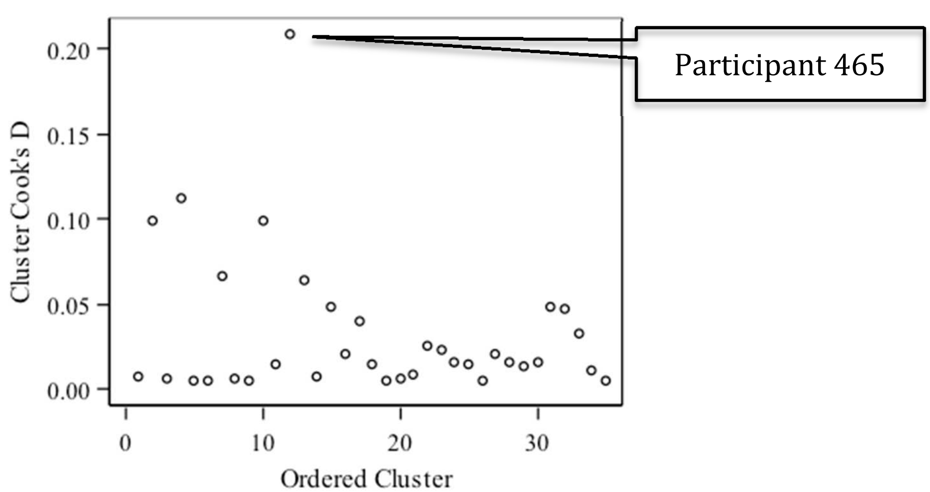
With regards to residuals, Cooks distance were graphed, both for individual records (Figure 6) as well as grouped by participant ID (Figure 7) to evaluate for statistical outliers that could unduly influence our model.

Figure 6: Cooks distance for individual records



Individual record outliers 699, 478, 621 and 623 (demonstrated above) were further explored, no data entry errors were found. The outlying records were excluded from the full GEE model in stepwise fashion with no significant change in associations or estimates.

Figure 7: Cooks distance grouped by participant.



Clustered Cooks Distance demonstrated one full participant ID #465, with possible excessive influence. No data entry errors were found. The outlying record ID was excluded from the full GEE model, again, with no significant change in overall correlations.

Table 9 below summarizes the original GEE full model with all records and participants, compared to models that: a) excluded 2 records with most influence, b) excluded 4 records with most influence c) excluded 1 whole participant with most influence.

Table 9: GEE full model compared with model excluding 2 records, 4 records and 1 participant.

	GEE Full Model			Excluding 2 records			Excluding 4 records			Excluding 1 participant		
Variable	OR	SE	p value	OR	SE	p value	OR	SE	p value	OR	SE	p value
Interruption	0.80	0.21	0.31	0.79	0.21	0.31	0.84	0.24	0.46	0.81	0.24	0.38
Position												
Intern (base)	1.00			1.00			1.00			1.00		
Senior Res	1.29	0.31	0.44	1.28	0.33	0.44	1.38	0.33	0.28	1.55	0.32	0.17
Attending	2.40	0.25	<0.01	2.32	0.27	<0.01	2.45	0.27	<0.01	2.50	0.27	<0.01
Type of ECG												
Normal (base)	1.00			1.00			1.00			1.00		
Anterior STEMI	0.78	0.49	0.61	1.00	0.61	0.98	1.45	0.61	0.54	1.14	0.52	0.8
Inferior STEMI	0.06	0.31	<0.01	0.06	0.34	<0.01	0.06	0.34	<0.01	0.06	0.34	<0.01
Scenario Exam	1.01	0.02	0.62	1.01	0.02	0.89	1.01	0.02	0.56	1.03	0.03	0.31
Confidence ECG	3.68	0.43	<0.01	3.62	0.25	<0.01	3.27	0.24	<0.01	3.43	0.25	<0.01

<u>Sensitivity Analysis</u>

Given the assigned dichotomous definition of confidence, a sensitivity analysis of the interaction effect of the dichotomous confidence score redefining a low score as Likert 1-2 and high score as Likert 3-5 was conducted. The interaction of confidence*interruption remained significant, (OR 15, p<0.01).

Discussion

The current study evaluated clinical interruptions in a simulated task-switching environment compared with an uninterrupted environment, focusing on accuracy of the interrupting task, ECG interpretation for STEMI. Interpretation of screening ECG's for STEMI represents a common interrupting task that requires immediate attention. STEMI is regarded as a medical emergency; delays in diagnosis result in increased morbidity and mortality (16-18). As such, guidelines recommend patients presenting with chest pain have a screening ECG for STEMI performed and interpreted by an emergency provider within 10 minutes of arrival to the ED. (19)

Studies evaluating emergency provider's accuracy in interpreting ECG's for STEMI show significant variability. (20-25) Our results, demonstrating that provider experience associates with improved accuracy of ECG interpretation for STEMI are consistent with prior literature. Further, ECG difficulty, as measured by both type of ECG (specifically inferior STEMI) and lower self-reported confidence in interpretation, also were associated with decreased accuracy of interpretation, consistent with past studies (26) However, contrary to our initial hypothesis, accuracy of ECG interpretation did not significantly change in a task switching environment compared to an uninterrupted environment, even after adjusting for these factors.

There are many possible reasons why an interrupted environment was not associated with ECG interpretation. First, it is possible that the ECG's selected for the clinical stimuli were not challenging enough. To avoid ambiguity in interpretation, ECG's were specifically selected based on published ACC/AHA guidelines for STEMI (19). Easier ECG's selected for interpretation may have limited the impact of the interruption on working memory, causing no measurable disruption in task performance. However, despite the selection of unambiguous ECG's, provider's accuracy, on average, was only 80%-90%, both in the pilot cohort as well as the main study population. Additionally, significant differences in accuracy were found in subgroups in which we expected a difference based on the literature, including by position (intern, senior resident, attending) and by type and difficulty of ECG.

Second, it is possible that the time to complete interpretation of interruptions, 15 seconds, was too long to overload working memory. The decision to pause for 15 seconds to allow evaluation and interpretation of clinical stimuli was based on subjective observations of emergency providers as well as a simulated practice session with author and experienced emergency physician TM. If the pause to interpret clinical interruptions was too long, it may have allowed the participant to engage in strategies, such as priming, to decrease the strain on working memory. However, our clinical experience during the simulation session found that many participants utilized 15 seconds or more to interpret the clinical stimuli.

Finally, it is possible that task switching workflow had less of an effect on errors than previously thought. It is suggested in prior literature that not all interruptions have the same effect on working memory. Interruptions that are longer, use similar cognitive resources, and occur in the middle of the primary task often lead to worse performance. (27-29) Our clinically based simulation utilized patient scenarios (auditory and visual cognitive modality) and short, interrupting clinical stimuli (primarily visual cognitive modality). It is possible that the length and nature of the interruption itself may have, in an emergency physician trained to interpret STEMI, only minimally stressed working memory, leading to no obvious difference in overall error rates with and without interruption.

However, when we examine the subgroup of difficult to interpret ECG's, as determined by providers confidence scores, we find interruptions decreased odds of correct interpretation, suggesting those ECG's that required increased cognitive resources to interpret suffered decreased accuracy in the task switching environment.

Overall, not all interrupted clinical tasks are prone to error. Specifically, those clinical stimuli where providers indicated high confidence in correct diagnosis had no significant difference in error rate, whereas those with low confidence, possibly utilizing more cognitive resources given their difficulty, has a significantly lower probability of correct interpretation. Our results suggest that providers may be able to determine which clinical interrupting tasks have a higher likelihood of error in an interrupting environment, based on their overall confidence in their errors.

Limitations

There are several limitations to the current study. First, our study size was restricted to the emergency providers at a single institution who volunteered to complete the study. However, our *a priori* sample size analysis estimated that 33 participants was likely sufficient to demonstrate a clinically significant difference in overall accuracy.

Second, participants were not randomized to order of interruption module (task-switching or uninterrupted). We chose to have all participants complete the task-switching module first and the uninterrupted group second to increase power (as compared to dividing participants into two groups of 17 people each), as well as to avoid discovery of the primary outcome during the uninterrupted module, which may have lead to participants devoting more time to the interrupting stimuli rather than balancing among all tasks. Further, we wanted to avoid any effect of priming by completing multiple ECG's immediately prior to the task-switching scenario. By not randomizing whether the interrupted module was given first or second, we hypothesized that we risked introducing type 1 error, possibly improving scores in the uninterrupted module after participants had already completed task switching. However, we designed our study to be externally valid, as the task switching was meant to simulate emergency workflow without prior preparation and the uninterrupted scenario simulated a test or exam. Further, by using a crossover study, the participant acted as their own control, reducing variability and bias.

Strengths

This is the first study to our knowledge, to evaluate clinically relevant interruptions in a simulated emergency environment. Further, we have created clinical stimuli and patient scenarios that may be utilized to easily replicate the study in other specialties and institutions. The study adds evidence that not all interruptions are prone to the same rate of clinical errors; providers may be able to self-screen complex clinical interruptions that may be more prone to error in a task-switching environment.

Conclusions

While there were differences in odds of correct ECG interpretation by position and by difficulty of ECG, we found no evidence overall that odds of correct ECG interpretation for STEMI significantly decreased in a task-switching environment. However, regarding difficult to interpret ECG's as determined by confidence scores, odds of correct ECG interpretation decreased in an interrupted setting. This suggests that odds of correct interpretation depend not only on participant level factors (including experience) but also on participant's confidence in interpreting stimuli. Further research should explore variables, both at the participant level and the interruption level, that increase odds of clinical error. By defining variables that increase odds of error, we may be able to focus efforts to mitigate the effects of high risk interruptions on medical errors.

Works Cited

1) Brixey JJ, Tang Z, Robinson DJ, et al. Interruptions in a level one trauma center: A case study. *International Journal of Medical Informatics*. 2008;77(4):235–241.

2) Chisholm CD, Dornfeld AM, Nelson DR. Work Interrupted: A Comparison of Workplace Interruptions in Emergency Departments and Primary Care Offices. Annals of Emergency Medicine, 2001;38(2): 146-151.

3) Spencer R, Coiera E, Logan P. Variation in communication loads on clinical staff in the emergency department. Annals of Emergency Medicine. 2004;44:268-273.

4) Chisholm CD, Weaver CS, WhenmouthL, Giles B. A task analysis of emergency physician activities in academic andcommunity settings. Annals of Emergency Medicine. 2011;58:117–122

5) Institute of Medicine. To err is human: building a safer health system. Washington DC: National Academy Press, 2000

6) Speier C, Valacich JS, Vessey I. The Influence of Task Interruption on Individual Decision Making: An Information Overload Perspective. Decision Sciences. 1999; 30(2) 337-360.

7) Altmann EM, Trafton JG. Time course of recovery from task interruption: Data and a model. Psychonomic Bulletin & Review 2007; 14 (6): 1079–1084

8) Dismukes K, Nowinski J. Chapter 16. Prospective Memory, Concurrent Task Management, and Pilot Error. Appears in Kramer A, Wiegmann D, Kirlik A, Attention: From Theory to Practice. 2007:1–12.

9) Trafton JG, Altmann EM, Brock DP, Mintz FE. Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*. 2003;58(5):583–603.

10)Magrabi F, Li SYW, Dunn AG et al. (2011) Challenges in Measuring the Impact of Interruptions on Patient Safety and Workflow Outcomes. Methods Inf Med. 50; 447-453.

11) Westbrook JI, Coiera E, Dunsmuir WT, et al. (2010) The impact of interruptions on clinical task completion. Qual Saf Health Care.19: 284-289.

12) Flynn EA, Barker KN, Gibson JT, et al. Impact of interruptions and distractions on dispensing errors in an ambulatory care pharmacy. Am J Health Syst Pharm 1999;56:1319–1325.

13) Magrabi F, Li SYW, ODay, R, et al (2010) Errors and electronic prescribing: a controlled laboratory study to examine task complexity and interruption effects. J Am Med Inform Assoc. 17:575-583

14) Monteiro SD, Sherbino JD, Ilgen JS et al. (xxx) Disrupting Diagnostic Reasoning: Do Interruptions, Instructions, and Experience Affect the Diagnostic Accuracy and Response Time of Residents and Emergency Physicians? Academic Medicine.

15) Goodell KH, Cao, CGL, Schwaitzberg SD (2006) Effects of Congnitive Distraction on Performance of Laparoscopic Surgical Tasks. Journal of Laparoendoscopic and Advanced Surgical Techniques. 16 (2) 94-98.

16) Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients. Fibrinolytic Therapy Trialists' (FTT) Collaborative Group. The Lancet. 1994;343(8893):311–322.

17) Udelson JE, Beshansky JR, Ballin DS, Feldman JA. Grif fi th JL, Heller GV, Hendel RC, Pope JH, Ruthazer R, Spiegler EJ, Woolard RH, Handler J, Selker HP. Myocardial perfusion imaging for evaluation and triage of patients with suspected acute cardiac ischemia. JAMA. 2002;288:2693–2700.

18) McCarthy BD, Beshansky JR, D'Agostino RB, Selker HP. Missed diagnoses of acute myocardial infarction in the emergency department: results from a multicenter study. YMEM. 1993;22(3):579–582.

19) Antman EM. ACC/AHA Guidelines for the Management of Patients With ST-Elevation Myocardial Infarction--Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 1999 Guidelines for the Management of Patients With Acute Myocardial Infarction). *Circulation*. 2004;110(5):588–636.

20) Brady WJ, Perron A, Ullman E. Errors in emergency physician interpretation of ST-segment elevation in emergency department chest pain patients. Acad Emerg Med. 2000;7(11):1256–1260.

21) RF S, JM R. Electrocardiogram interpretation as a basis for thrombolysis. J R Coll Physicians Lond. 1997;31(5):536–540

22) Snoey ER, Housset B, Guyon P, et al. Analysis of emergency department interpretation of electrocardiograms. Journal of accident & emergency medicine. 1994;11(3):149–153.

23) Kuhn M, Morgan MT, Hoffman JR. Quality assurance in the emergency department: evaluation of the ECG review process. YMEM. 1992;21(1):10–15.

24) Tricomi AJ, Magid DJ, Rumsfeld JS, et al. Missed opportunities for reperfusion therapy for ST-segment elevation myocardial infarction: Results of the Emergency Department Quality in Myocardial Infarction (EDQMI) study. American Heart Journal. 2008;155(3):471–477.

25) Armstrong EJ. Prevalence and Factors Associated With False-Positive ST-Segment Elevation Myocardial Infarction Diagnoses at Primary Percutaneous Coronary Intervention–Capable Centers. Arch Intern Med. 2012;172(11):864.

26)Brady WJ, Perron AD, Chan T. Electrocardiographic ST-segment elevation: correct identification of acute myocardial infarction (AMI) and non-AMI syndromes by emergency physicians. *Acad Emerg Med*. 2001;8(4):349–360.

27) Rivera-Rodriguez AJ, Karsh BT. Interruptions and distractions in healthcare: review and reappraisal. Qual Saf Health Care 2010;19:304-312

28) Gillie T, Broadbent D. What makes interruptions disruptiveda study of length, similarity, and complexity. Psychol Res 1989;50:243-250.

29) Edwards MB, Gronlund SD. Task Interruption and its Effects on Memory. Memory 1998;6(6):665-687.