AHRQ Grant Final Progress Report

Title of Project:

Analysis of Factors Associated with Clinical Checklist Compliance

Principal Investigator:

Aleksandra Sarcevic, PhD
Associate Professor, College of Computing and Informatics, Drexel University

Team Members:

Ivan Marsic, PhD
Professor, School of Engineering, Rutgers University

Randall S. Burd, MD, PhD
Division Chief, Trauma and Burn Surgery

Leah Hamlin, PhD
Graduate Research Assistant, Drexel University

Organization:

Drexel University

Inclusive Dates of Project:

09/30/2018 – 09/29/2020

Federal Project Officer:

Emily Chew

Acknowledgment of Agency Support:

This project was supported by grant number 5R03HS026057-02 from the Agency for Healthcare Research and Quality.

Grant Award Number:

5R03HS026057-02

Structured Abstract

Purpose: The goal of this project was to identify context attributes associated with noncompliant use of a checklist that is administered concurrently with clinical activities being performed during trauma resuscitation.

Scope: Checklists may be feasible for concurrent use in complex and dynamic clinical settings, but compliance with checklists has been suboptimal. Prior research identified barriers to checklist compliance but did not examine the effects of task complexity and different contexts on checklist compliance.

Methods: The project analyzed 187 pediatric trauma resuscitations, in which team leaders used the digital checklist to guide patient care. For each case, the project obtained digital checklist logs, data on about 10 context attributes, and activity logs. Measures of compliance included the number of unchecked items, number of checklist notes, checklist completion in relation to task performance, and timing of check-offs.

Results: Four contexts affected checking and note-taking behaviors on the digital checklist: (1) team leader experience level, (2) team leader's arrival to the resuscitation room after the patient, (3) patients with a penetrating injury, and (4) patients with external injuries. The project also identified three noncompliant behaviors: (1) false checks - item checked but the corresponding task not performed, accounting for 16% of all checks; (2) inaccurate checks - item checked and the corresponding task started but not performed to completion, accounting for 5% of all checks; and (3) failed checks - item not checked but the corresponding task performed to completion, accounting for 49% of all unchecked items.

Key Words: checklist compliance, digital checklist, trauma resuscitation

Purpose

The goal of this project was to identify context attributes associated with noncompliant use of a checklist that is administered concurrently with clinical activities being performed during pediatric trauma resuscitation. Our central hypothesis was that context attributes impact the likelihood of checklist compliance. To test this hypothesis, we pursued two specific aims:

Specific Aim #1: Identify patient, provider, task, and event attributes that are associated with clinical checklist compliance. Our *working Aim 1 hypothesis* was that patient attributes, provider experience, task duration, and time of day influence checklist compliance.

Specific Aim #2: Determine how task complexity affects the timing and accuracy of interactions with a clinical checklist. Our working Aim 2 hypothesis was that noncompliant checklist use is more common for tasks that are multi-step, are parallel, and have extended duration.

Scope

Background

Most medical checklists have been developed for use either before initiating or after concluding clinical tasks, when sufficient time is available for completing the checklist. There are, however, several areas of healthcare in which complex situations require rapid approaches to patient management (e.g., emergency medicine and anesthesiology), making this standard model of checklist use impractical. Recent work has suggested that checklists may be feasible for concurrent use in these complex and dynamic clinical settings, when compliance with protocols and timely action are essential. In simulated operating room crisis scenarios, the concurrent use of a checklist was associated with improved team performance and adherence to critical steps in management.[1] Our research group similarly showed that the use of an Advanced Trauma Life Support (ATLS) resuscitation checklist improved the completeness and timeliness of critical evaluation steps, particularly for patients arriving without prior notification.[2]

Not all checklist implementations result in improved team performance and patient safety. Multiple studies have shown negative effects of checklists on medical work, attributing those effects to low checklist compliance due to lack of user training or poor checklist design.[3, 4] Checklist compliance has been defined as the number of completed sections or items,[5] as a combination of completion and accuracy,[6] and by whether or not the checklist was actually used.[7] To understand compliance, most prior studies used pre- and post-implementation design, collecting data through surveys and retrospective analyses.[7-10] Checklist success measures have mostly focused on adherence to protocols, mortality rates, and infection or complication rates. Although this prior research identified barriers to checklist compliance (e.g., poor team communication, issues with the checklist interface), few studies have looked at actual interactions with checklists and how task complexity and different patient and case features affect checklist use.

In this project, we focused on identifying the factors and underlying causes that contribute to noncompliant use of a concurrent checklist during pediatric trauma resuscitations at Children's National Health System's level-1 trauma center. To achieve this goal, we deployed a digital checklist for pediatric trauma resuscitation that we previously designed,[11,12] and we studied its use during live resuscitations occurring between January 2017 and April 2018.

Context

Our research team designed and implemented a digital version of the 55-item paper checklist for trauma resuscitation in October 2016.[11,12] This digital checklist application mirrors the design of its paper counterpart, with several feature enhancements afforded by tablet computers. The application includes the same five sections, separated into tabs: pre-arrival plan, primary survey

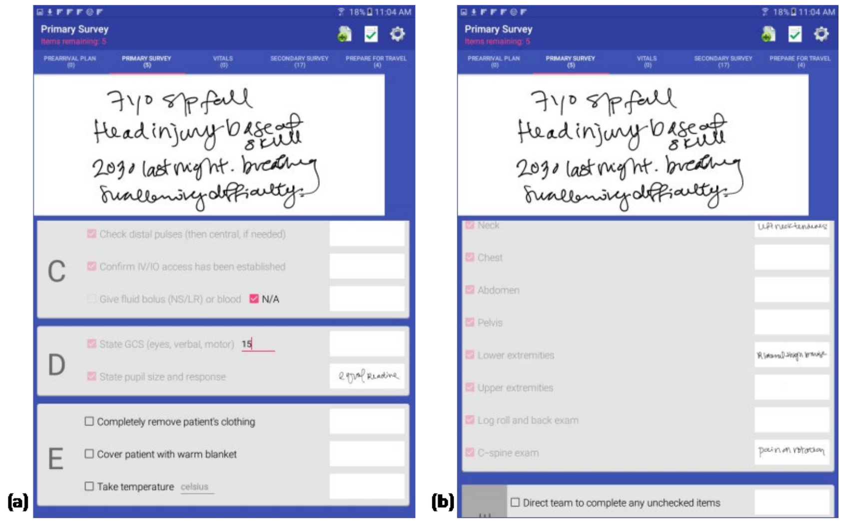

Figure 1: Example screens from the digital checklist with user notes in the margin area, typed and stylus notes on the primary survey (a), and checks/notes on the secondary survey (b).

and vital signs, secondary survey, and departure plan (Figure 1). A built-in tablet stylus is available for taking notes in the margin space or in note fields associated with items. Numerical items like patient weight or temperature have text entry fields for typing in values. As the leader is checking off items, the timestamp and corresponding item are saved to a checklist log. At the completion of the resuscitation, the leaders “submit” their checklists, triggering a review screen that shows any unchecked items. Users can check the remaining items using the review screen or can go back to the checklist tabs before completing the checklist and submitting the log. In addition to the list of checked items and timestamps (e.g., “01:31:24, Confirm airway is protected”), the checklist log includes values from typed notes, handwritten notes, any items left unchecked, and tab switching sequences.

This digital checklist has been in use during resuscitations at Children’s National Health System’s level-1 trauma center since October 2016. Upon arriving to the trauma bay, the leaders can choose between a paper or digital checklist format. Although administering a checklist is mandatory at the hospital, in some cases leaders skip using it due to their preferences or urgency of the situation. The leaders typically start using the checklist as the team prepares for patient arrival and later performs exams and treatments. The concurrent checklist administration is achieved by calling out items and waiting for responses from team members who are performing related tasks before checking them off. Some leaders rely on verbal reports that signal task completion or on their own observations of team activities before they check off items on the checklist.

Research Setting

Children’s National Medical Center (CNMC) in Washington, DC, is a 280-bed acute care hospital with a level-1 trauma center that treats about 1,000 injured children each year. Over 600 patients are initially treated in one of two adjoining resuscitation rooms in the emergency department. The resuscitation rooms are equipped with a video and audio recording system with two cameras (above the patient and wide-angle) and three microphones for capturing live resuscitations. The data are recorded to a dedicated server that is accessible via a password-protected portal. Data from the patient’s record are entered into a clinical database by a dedicated data registrar and are used for benchmarking, performance improvement, and research purposes. The project was approved by the hospital’s Institutional Review Board (IRB), and a Reliance Agreement between the Drexel University and CNMC IRBs was obtained, with the hospital IRB serving as the IRB of Record.

Participants

A total of 32 trauma team leaders participated in this project as checklist users. The leaders included 20 senior surgery residents and 12 surgery fellows, with trauma resuscitation experience ranging from 1 month to 7 years. The leaders used the checklist an average of 12 times each (SD: 18; median: 5). All team leaders consented to participating in the research study and were trained on the system before they started using it for patient care. Participants received the IRB information sheet that described the study details, data confidentiality, and any risks and benefits associated with participating in the study.

Dataset

Between January 2017 and March 2018, the center treated 517 trauma patients. Of these, 431 cases had a signed consent and video files available for research purposes. The paper checklist was used in 150 resuscitations, and the digital checklist was used in 187. Ninety-three resuscitations were performed without a checklist. Surgery fellows led 47.3% of resuscitations, senior residents led 49.9%, and 2.8% of resuscitations were performed without a surgical team leader. The leader was present before the patient arrived 69% of the time. Most cases (60.6%) were triaged as stat activations, 26.5% were transfers, and 13% were attending activations; similar distributions were observed for both digital and paper checklist cases. "Now" activations occurred in 9% of digital checklist cases. Blunt was the most frequent injury mechanism (91.2%). Most patients had an external injury (57.8%), and most reported a GCS of 14 or 15 (84.9%). We observed two significant differences in patient and provider attributes between all checklist and no checklist resuscitations: (1) team leader presence at the time of the patient arrival was associated with checklist use ($p<.001$), and (2) "now" activations were associated with no checklist use ($p<.001$). These results confirmed that the checklist was not used when the surgical team leader was absent and when the patient arrived before the team. For all other contexts, no statistically significant differences were found between cases with and without the checklist. In addition, no significant differences were observed between digital checklist and paper checklist cases, showing that our data were not biased toward any particular patient or case attribute.

Methods

Specific Aim 1

Study Design: Our working hypothesis was that a discrete set of contextual elements, such as patient attributes, provider experience, task duration, and time of day, influences checklist compliance. For each of the 187 digital checklist cases collected for this project, we obtained 20 attributes related to patients, providers, and environment from the hospital's trauma registry and medical records. This information was de-identified and included gender, age (years and months), whether the patient was intubated, the leader's experience level, whether the leader was present when the patient arrived, whether the case was a "now" activation, the activation level, the mechanism of injury, the injured body region, the injury severity score (ISS), the patient's neurological status (Glasgow Coma Score or GCS), GCS eye score, GCS verbal score, GCS motor score, the Emergency Department disposition, final disposition, length of stay (LOS), weekday or weekend event, day or night-time event, and the number of times each participant used the digital checklist over the study period. From these 20 attributes, we selected 10 that were independent from other attributes after we ran the Fisher's exact test and that were more likely to affect concurrent use of the checklist during trauma resuscitation. The final 10 attributes were (1) trauma team leader experience level, (2) trauma team leader presence at the time of patient arrival, (3) now activations, (4) team activation level, (5) mechanism of injury, (6) injured body region (AIS score), (7) neurological status (GCS score), (8) time of the day, (9) time of the week, and (10) frequency of checklist use.

Measures: We determined the measures of checklist use based on prior research that found the importance of note taking for memory work[13] and high completion rates for improved patient outcomes.[1] Our two main measures included the number of unchecked items and the number of notes on the checklist. We added four measures that were specific to note taking: the number of margin notes, the number of item-associated notes, the number of numerical notes, and the number of narrative notes.

Data Collection & Analysis: For each of the 187 cases, we collected digital checklist logs. The logs included the following information: checked items and timestamps for those checks (e.g., “22:45:15 Log roll and back exam”), values from typed notes (e.g., “Blood pressure: 132/84”), handwritten notes and corresponding checklist items, handwritten margin notes, and any unchecked items. We transcribed notes from the logs and classified them based on location of note (margin or item), type of note (narrative or numerical), length of note (one word, two to four words, or five or more words), and category of note (exam finding, task status, pre-hospital information, or care plan). Log transcripts were then matched with context attributes for each case using the assigned case numbers. We used multivariate regression as our primary analysis to identify associations between the study contexts and our two main checklist metrics—number of total notes and number of unchecked items. We selected the Mann-Whitney U test to compare each of the 10 contexts with each of the six measures, because this test does not require the assumption of normal distribution. Because our univariate analysis included a large number of comparisons (nine context variables times six checklist variables), we used a Bonferroni correction to reduce the likelihood of a type I error (adjusted p values to $p<.001$). Finally, we ran a linear regression with the team leader checklist use frequency as an independent variable to determine if note taking or checking behaviors changed with the increased use of the system.

Limitations: We focused on 10 of 20 context attributes that we determined would most likely impact the checklist use. Some of the selected contexts included only a few cases (e.g., penetrating injuries were present in seven of 187 cases). Though these numbers were small, they were representative of the overall number and types of cases seen at the hospital during the data collection period.

Specific Aim 2

Study Design: Our working hypothesis was that noncompliant checklist use is more common for tasks that are multi-step, are parallel, and have extended duration. We first determined the types of noncompliant checking behaviors and task features contributing to those behaviors. We then analyzed how the timing of checking items correlated to completion times of the corresponding tasks. From the 187 cases collected for this project, we excluded cases that did not have the ground truth activity data and checklist log data available, leaving us with 98 cases for this Aim.

Measures: Our measure of compliance was checklist completion in relation to task performance, because we studied the extent to which checklist completion represents actual work processes. We also added timing of check-offs as a new measure, given our focus on concurrent use of checklists during time-critical task performance.

Data Collection & Analysis: Medical experts at the research site used a video annotation software to review videos from three camera angles and to code and timestamp task performance in all 98 cases. The final activity log for each case includes task start and finish times, the team role performing the task, and definitions for task completion (e.g., whether the activity was verbalized, incomplete, or performed to completion). From these activity logs, we extracted data for 32 tasks corresponding to specific checklist items. We selected 32 (of 55) checklist items based on their clinical relevance and type (e.g., assessment vs. treatment tasks): three tasks from the pre-arrival plan section, 11 from the primary survey, four from vital sign items, and 14 items from the secondary survey section.

Table 1: *Specific Aim 1 Findings*: Associations between patient, provider, and environment attributes, and digital checklist metrics. Asterisks indicate significant p values after Bonferroni correction at <.001.

Attributes	Median# Notes (IQR)	*p* value	Mean # Margin Notes (25%,75%)	*p* value	Median # Item Notes (IQR)	*p* value	Median # Narrative Notes (IQR)	*p* value	Median # Numerical Notes (IQR)	*p* value	Median # Unchecked Items (IQR)	*p* value
TL type		**<.001***		0.003		0.303		0.789		**<.001***		0.051
Fellow	8(7,11)		0.25(0,1)						7(6-7)		2(0-6)	
Senior res	6(2,8)		0.46(0,1)						4(0-6)		0(0-7)	
TL at pt. arrival		0.51		0.34		0.076		0.36		0.027		0.003
Present									6(2-7)		1(0-5)	
Not Present									4(0-7)		7(0-10)	
Activation level		0.34		0.68		0.18		0.54		0.04		0.71
Attending									2(0-7)			
Not attend.									6(2-7)			
Now activation		0.12		0.962		0.018		0.085		0.155		0.005
Now					0(0,1)						7(1,9)	
Not now					0(0,11)						1(0,6)	
Injury mechanism		0.08		**<.001***		0.78		0.42		0.004		0.034
Blunt			0.3(0,1)						6(2-7)		1(0-6)	
Penetrating			0.8(0,1)						1(0-4)		6(0-12)	
GCS total		0.23		0.32		0.059		0.006		0.52		0.61
9-13							0(0-1)					
3-8, 14-15							1(0-3)					
Inj. body region												
Head/Neck/Face		0.38		0.28		0.19		0.24		0.75		0.79
Chest/Abd/Pelv.		0.79		0.18		0.82		0.65		0.49		0.62
Extremities		0.11		0.31		0.009		0.008		0.68		0.84
No injury					1(0-3)		0(1-3)					
Injury					0(0-2)		0(0-1)					
External		0.014		0.053		**<.001***		**<.001***		0.92		0.13
No injury					1(0-1.5)		1(0-2)					
Injury					1(1-4)		2(1-4)					
Day of week		0.23		0.304		0.93		0.53		0.04		0.66
Weekday									6(1-7)			
Weekend									6(4-7)			
Time of day		0.59		0.402		0.27	0.11			0.018		0.64
Day									5(1-7)			
Night									7(1-7)			

Based on the item to task matching criteria, we noted whether each task was performed to completion, started but not completed (incomplete), or not performed at all. For each task that was performed to completion, we noted whether or not the checklist item was checked. When the task was missing from the activity log (i.e., not performed at all), we also noted whether or not the checklist item was checked. Finally, when the task was started but not performed to completion, we noted whether or not the checklist item was checked. We calculated checklist compliance for each of the 32 tasks in all 98 cases.

We wrote a Java script to extract and parse all timestamps and labels for 32 checklist items from the checklist logs as well as corresponding task labels and performance times from the activity logs. Because checklist users are trained to check off boxes only after the first instance of tasks that are performed multiple times (e.g., blood pressure measurement), we extracted only the first instances of this type of task. We then compared the timestamps for checklist items and corresponding tasks, expecting three possible timepoints when check-offs occurred in relation to task performance: before, during, or after task performance. Finally, we reviewed videos of all 98 cases to further understand factors contributing to noncompliant checklist use.

Limitations: Not all checklist items were selected for analysis, because we focused on time-critical tasks. This was a single-site study, and checklist use could differ at other centers.

Results

Principal Findings and Outcomes – Specific Aim 1

By comparing six checklist metrics across 10 contexts defined by patient, provider, and environment attributes, we identified four major contexts that affected checking and note-taking behaviors on the digital checklist: (1) team leader experience level, (2) team leader's arrival to the resuscitation room after the patient, (3) patients with a penetrating injury, and (4) patients with external injuries (Table 1).

Team Leader Experience Level: The multivariate and univariate regression analyses showed that fellows (more experienced leaders) took significantly more notes than senior residents ($p<.001$ for both analyses, Table 1). Among the 187 checklists, 94 were completed by fellows and 93 by senior residents. These results showed that fellows used the checklist more frequently, taking a median of eight notes per checklist, whereas senior residents took a median of six notes. When it came to margin notes, however, we found that senior residents (less experienced leaders) took more notes in the checklist margin than did fellows ($p=0.003$, Table 1). Among the 187 digital checklists, 92 contained margin notes. We observed two trends in leaders' margin notes. First, each of the 92 cases had a handwritten margin note about the reported pre-hospital information, including patient age and mechanism of injury (e.g., "18m/o fall down concrete steps"). Second, most margin notes were extensive, containing five or more words. We also found that fellows took significantly more numerical notes than senior residents did ($p<.001$, Table 1). The median number of numerical notes per checklist taken by fellows was seven (Table 1), showing that fellows entered notes for almost all numerical fields (out of possible eight per case): patient weight, Glasgow Coma Score, pupils, temperature, and four vital signs (heart rate, respiratory rate, oxygen saturation and blood pressure).

Trauma Team Leader Presence at the Time of Patient Arrival: Multivariate regression analysis showed that cases with the leader absent at patient arrival had significantly more items left unchecked, a median of seven per checklist ($p=0.018$, Table 1). The leader arrived late (after the patient) in 29 of 187 cases. The most frequently unchecked items were in the pre-arrival plan (115 times left unchecked in 29 cases, an average of four pre-arrival plan items per checklist) and Pause sections (44 times in 29 cases, an average of 1.5 pause items per checklist). Seventeen of the late-arrival cases were led by fellows and 12 by senior residents.

Mechanism of Injury: We found significant differences in how the checklist was used between patients with penetrating injuries and those with a blunt injury. These differences were observed in the number of margin notes and in the number of unchecked items. For example, we observed a higher frequency of margin notes for patients with penetrating injuries ($p<0.001$, Table 1). Multivariate regression analysis also showed that patients with a penetrating injury had more unchecked items ($p<.001$, Table 1) than patients with blunt and other injury types. In the seven penetrating injury cases, the frequency of unchecked items was consistent across checklist sections: 17 total unchecked in the pre-arrival plan section, 15 unchecked in pause sections, 14 unchecked in the secondary survey section, 11 unchecked in the primary survey section, and five unchecked in the prepare for travel section. These findings show that completion rates decrease in cases with highly acute patients.

Table 2: *Specific Aim 2 Findings*: Frequency of false, inaccurate, and failed checks for each of the 32 tasks.

Task	**False Checks (n=452)**	**Inaccurate Checks (n=154)**	**Failed Checks (n=134)**
Pre-arrival Plan			
Oxygen Equip.	6	0	14
Suction Equip.	8	0	11
Bair Hugger	6	1	9
Primary Survey			
Airway assessment	2	14	2
C-spine stabilized	10	0	9
Confirm O_2	20	0	4
IV/IO Access	48	5	4
Fluid/blood	0	0	10
Pulses	1	38	0
GCS verbalized	7	1	2
Pupils	3	0	3
Remove Clothing	27	10	4
Warm Blanket	19	0	7
Temperature	3	0	15
Heart Rate	11	0	3
Respiratory Rate	9	0	4
Oxygen Saturation	8	0	3
Blood Pressure	0	0	2
Secondary Survey			
Head	6	0	1
Ears	9	5	2
Ocular Integrity	94	0	0
Facial Bones	9	0	3
Nose	39	18	3
Mouth	22	4	1
Neck	33	32	2
Chest	1	1	2
Abdomen	0	0	2
Pelvis	13	1	1
Lower extremities	4	8	2
Upper extremities	13	16	1
Back exam	4	0	3
C-spine exam	17	0	5

Injured Body Region: Of the four injured body regions, checklist use was significantly different only for patients with external injuries (AIS6). Of 187 patients, 111 (59.4%) had external injuries. Our results showed that patients with external injuries had significantly more item notes than patients without external injuries ($p<.001$, Table 1). Patients with external injuries also had significantly more narrative notes ($p<.001$, Table 1), an expected finding given their descriptive function (narrative notes include both margin and item-associated notes).

Discussion – Specific Aim 1

Most of the unchecked items in the late arrival scenarios were found in the pre-arrival section. These tasks are critical for team preparation, but their omission did not impact the completion of other checklist tasks. Even so, late arrival of any team member, especially of a team leader, adds to the chaotic nature of the environment. Approaches to checklist completion that are less reliant on user interactions, such as using sensor data to automatically detect completed preparatory tasks may be feasible in these situations.

The distribution of unchecked items for patients with penetrating injuries is more challenging to address. Nonroutine cases, such as those with patients in extremis, would most benefit from cognitive aids due to a higher likelihood of errors and omitted steps. Although our data showed that leaders took more notes in those situations, they were also less compliant with the checklist (i.e., leaving many items unchecked). Existing checklists are currently showing nonroutine tasks along the other routine tasks, precluding providers from focusing on those critical tasks at the time they need them. The dynamic listing, ordering, and highlighting of the checklist items at the time when they are needed would better support this complex workflow. The challenge, however, is in detecting those scenarios and recognizing the need for specific items based on current activity in the environment.

Conclusions – Specific Aim 1

Existing studies of concurrent checklist use in medical settings have focused on metrics such as team member location and have evaluated the system based on team performance rather than on individual interactions with the checklist. In emergency medical work, new contexts and information are emerging throughout the events, requiring technologies to rapidly absorb this information and adapt to users in meaningful ways. Our goal in Aim 1 was to understand how these contexts affected user interactions with the digital checklist that was concurrently administered while clinical tasks were being performed. The results showed that concurrent checklists are feasible for this highly complex medical work but would benefit from adaptive features for more efficient and complete use. We found that two contexts significantly affected checklist completion rates: (1) Team leaders arriving after the patient and (2) patients with penetrating injuries were both associated with more unchecked items. We also found that three contexts significantly affected the number of notes recorded on the checklist: (1) patients with penetrating injuries, (2) patients with external injuries, and (3) team leader experience level.

Principal Findings and Outcomes – Specific Aim 2

Checklist Compliance Behaviors: We identified five checklist compliance outcomes. *True checks* (item was checked and corresponding task was performed to completion) and *true nonchecks* (item was not checked and corresponding task was not performed) were considered measures of a compliant behavior, because they accurately documented and reflected task performance during resuscitations. The noncompliant behaviors included (Table 2): *false checks* (item was checked but the corresponding task was not performed), *inaccurate checks* (item was checked and the corresponding task was started but not performed to completion), and *failed checks* (item was not checked but the corresponding task was performed to completion). Of the possible 3,136 checklist items (all 32 items in each of the 98 cases), we found a total of 2,862 check-offs (91%) and 274 unchecked items (9%). Of the checked items, 79% were true checks, 16% were false checks, and 5% were inaccurate checks. Of the unchecked items, 51% were true nonchecks, and 49% were failed checks.

<u>False checks</u> occurred for 29 of the 32 tasks and accounted for 16% of all checked items. The most frequently false-checked items included four on the primary survey and six on the secondary survey (Table 2). After reviewing videos for team and leader behaviors, we identified several patterns that explain these false, noncompliant checks-offs. For the six falsely checked items in the secondary survey, we found that their corresponding tasks were either performed as part of an already completed primary survey task or were grouped into an exam, comprising several individual tasks. For example, the ocular integrity exam is often completed as part of the primary survey step D, while the bedside physician is assessing pupils. Teams then skip this item when they reach the secondary survey, but physician leaders still check it off at this time, rather than at the time it was actually performed. In contrast, the experts who coded ground truth activity logs by following clinical practice guidelines marked this task as not performed, because it was not included in the secondary survey. For the four primary survey tasks, we observed two factors contributing to false checks: (1) tasks were performed incorrectly (and were therefore coded as not performed), but the leaders checked them off because they saw the team executing these items, or (2) tasks were performed before the patients arrived to the hospital (e.g., by emergency medical services teams transporting the patient), but the leaders checked them off regardless.

<u>Inaccurate checks</u> accounted for 5% of all checked items, occurring for 14 of the 32 tasks (Table 2). Through video review, we found that inaccurate checks mostly occurred when team leaders checked off items for tasks that were started but not completed. For example, we observed that the bedside physician assessed pulses on the lower extremities but skipped upper extremities (as required) because another team member was taking blood pressure or placing an IV on the upper extremities at the same time. Another factor contributing to inaccurate checks was the multistep nature of the task. Our analysis of five inaccurate checks for the “Confirm IV/IO access has been established” showed that team leaders prematurely checked off the item when they observed nurses starting this task, rather than waiting for confirmation that the task was completed.

Failed checks were found for 30 of the 32 tasks, accounting for 7% of all checks (a total of 134 failed across all 98 cases) (Table 2). The leaders failed to check off the completed tasks because they missed the verbalizations of the results due to noise.

Timing of Check-Offs in Relation to Task Performance: For an in-depth understanding of checklist interactions, it was critical to not only examine when did users check off items in relation to task performance but also what factors contributed to the variable timing.

Items Checked Before Task Start (Pre-Checks). We defined a pre-check as the check-off timestamp occurring before the start time of the corresponding task. Of all true-checks, 15% occurred before the team started performing a task. The most commonly pre-checked items included “Heart Rate,” “Back” exam, “Cover patient with warm blanket,” “Pelvis” exam, “Lower extremities” exam, “Confirm IV/IO access,” “State pupils size and response,” “State GCS,” “Take temperature,” and “Confirm C-spine is immobilized properly.” Our video review of leader and team behaviors showed that many of these tasks are multistep processes and take longer to perform.

Items Checked During Task Performance (Real-Time Checks). Sixteen percent of all true checks occurred as the corresponding tasks were being performed. Items with most frequent checks during task performance included “Confirm C-spine is immobilized properly,” “Back” exam, “Confirm O2 placement,” “Abdomen” exam, “Check distal pulses,” “State pupils size and response,” and “State GCS.” Similar to pre-checked tasks, tasks checked off during performance are often longer in duration and may be performed more than once.

Items Checked After Task Completion (Post-Checks). Post-check occurred when the check-off timestamp occurred after the end time of the corresponding task performance. Post-checks were most common and accounted for 69% of all true checks, which is appropriate for a typical checklist use—tasks are first completed and then checked off.

Although post-task checking is typical for checklists, it was important to also determine the point at which post-checks became delayed, noncompliant behaviors (i.e., whether items were checked within a reasonable time period [slightly after task completion] or were checked long after task completion). We therefore calculated the median task duration (median time it took the team to perform a task from start to finish) and the median check-off delay (the time between the task ended to the moment the item was checked). For each task, we excluded pre-checks (noncompliant behaviors) and then analyzed the distribution of items that were checked off during and after task performance. The items with the longest median delay time included all pre-arrival plan items, all vital sign items, the warm blanket placement item on the primary survey, and C spine and Neck exams on the secondary survey. We considered these check-offs delayed, because it took the leaders close to 2 minutes or longer to check off the items.

Discussion – Specific Aim 2

Our analyses have shown that the three noncompliant checklist behaviors were not random but rather were caused by two major factors: (1) work practices and task perceptions that have formed at the bedside over time, and (2) the variable nature of task length and complexity.

The on-the-ground work practices, it turned out, differed from those recommended by clinical practice guidelines, as reflected through expert coding of task performance. The order of items on the checklist and their labels were derived based on established protocols and clinical practice guidelines. Resuscitation cases, however, vary based on the patient status, available team members, and other contextual factors. As we observed through video, trauma teams do not always follow the prescribed order of tasks during actual patient care. In actual practice, activities run continuously, stop and resume, overlap and intertwine.

The task length and complexity also affected how users interacted with the checklist. Our study found that the majority of tasks were checked off when their corresponding tasks were performed (79% true checks). Of those, 69% were checked post-task completion, but some check-offs occurred with long delays due to their complexity (e.g., multistep). Although our noncompliant results appear insignificant (e.g., 15% of items were pre-checked), even the smallest issues can have catastrophic consequences in settings where checklist use is considered one of the main safety procedures.

Conclusions – Specific Aim 2

Checklists have been shown to decrease human error in a range of settings, but little is known about the nature of user interactions with checklists. In this Aim, we studied the extent to which user interactions with a digital checklist represent actual activity in an emergency medical setting. We identified a set of noncompliant checklist behaviors and unpacked the factors contributing to noncompliant checklist use. To assist complex work processes, new checklist designs should consider task attributes and contexts within which those tasks are performed. Specifically, new checklist should better support performance of tasks that are performed as a group, tasks with long durations, tasks that are performed more than once, tasks with multiple steps, and tasks that are rarely performed. In addition, new checklist designs should also support retrospective checking behaviors.

Significance

This project addressed several important gaps in our knowledge about checklist compliance behaviors, especially in emergency medical environments, where checklists are administered concurrently with clinical tasks. First, the study identified current use practices for the checklist as a compliance tool and characterized how its use varies with different patient, provider, and environment attributes. Second, the project defined the relationships between checklist compliance and actual task performance through five checklist compliance outcomes: true checks, true nonchecks, false checks, inaccurate checks, and failed checks. The occurrence of false checking in particular poses risks to patient safety, because it provides false reassurance that clinical tasks have been performed. Understanding the causation of noncompliant checklist use is a required first step for improving the design and implementation of clinical checklists. Finally, through detailed analysis of checklist use behaviors and the nature of associated tasks, the project was able to determine how different task attributes contributed to noncompliant checklist behaviors.

Although this project was conducted in the setting of pediatric trauma resuscitation, our approaches and findings are generalizable to other medical domains that require rapid approaches to patient management, including the operating room, critical care units, and other resuscitation areas. The resuscitation room was an ideal setting for studying the relationship between compliance with concurrent checklists and actual task performance, because 1) trauma resuscitation is performed in a single site, enabling stationary monitoring; 2) the resuscitation protocol is standardized, facilitating systematic analysis of tasks completions and deviations; 3) errors of omission are frequent; and, 4) trauma resuscitation requires frequent critical medical decisions (approximately one every 72 seconds).[14]

Implications

Our unique dataset with time-stamped checklist interactions and activity logs allowed us to go beyond simple completion rates and understand how different contexts and the nature of tasks affected checklist use behaviors and checklist compliance. To assist complex work processes, new checklist designs should consider task attributes and contexts within which those tasks are performed. Our findings showed that activity performance cannot always be inferred based on user input on the checklist (i.e., checking off items), because checking is sometimes delayed or doesn't occur for some items. Rather, to adapt the checklist content for critical scenarios or to alert users of skipped tasks, the checklist system should combine user input with data from other sensors or sources of information in the environment. Similar approaches have been successfully explored in surgical settings, where motion and light intensity sensors were used to capture ventilator movement during surgery, which indicated patient intubation or extubation.[15]

Using the findings from this project, we can now advocate for an approach to checklist design that focuses on types of tasks, where the design will more accurately reflect the work "as is."

List of Publications and Products

1. Kulp L, Sarcevic A, Zheng Y, Cheng M, Alberto E, Burd RS. Checklist design reconsidered: Understanding checklist compliance and timing of interactions. Proc SIGCHI Conf Hum Factor Comput Syst. 2020 Apr;2020:10.1145/3313831.3376853. doi: 10.1145/3313831.3376853. PMID: 32685940; PMCID: PMC7368994.
2. Mastrianni A, Kulp L, Mapelli E, Sarcevic A. Understanding digital checklist use through team communication. Ext Abstr Hum Factors Computing Syst. 2020 Apr;2020:10.1145/3334480.3382817. doi: 10.1145/3334480.3382817. PMID: 32747878; PMCID: PMC7397407.
3. Kulp L, Sarcevic A, Cheng M, Burd RS. Towards dynamic checklists: Understanding contexts of use and deriving requirements for context-driven adaptation. ACM Trans Comput Hum Interact. 2021, in press.

In Preparation:

4. Sarcevic A, Kulp L, Alberto E, Ranganna A, Marsic I, Burd RS. Associations between activity attributes and checklist compliance in a time-critical medical setting. In preparation for submission to JAMIA.

References

1. Arriaga AF, Bader AM, Wong JM, Lipsitz SR, Berry WR, Ziewacz JE, Hepner DL, Boorman DJ, Pozner CN, Smink DS, Gawande AA. Simulation-based trial of surgical-crisis checklists. N Engl J Med. 2013 Jan 17;368(3):246-53. doi: 10.1056/NEJMsa1204720
2. Kelleher DC, Carter EA, Waterhouse LJ, Parsons SE, Fritzeen JL, Burd RS. Effect of a checklist on advanced trauma life support task performance during pediatric trauma resuscitation. Acad Emerg Med. 2014 Oct;21(10):1129-34. doi: 10.1111/acem.12487
3. Urbach DR, Govindarajan A, Saskin R, Wilton AS, Baxter NN. Introduction of surgical safety checklists in Ontario, Canada. N Engl J Med. 2014 Mar 13;370(11):1029-38. doi: 10.1056/NEJMsa1308261
4. de Vries EN, Prins HA, Crolla RM, den Outer AJ, van Andel G, van Helden SH, et al. Effect of a comprehensive surgical safety system on patient outcomes. N Engl J Med 2010;363(20):1928-37. doi: 10.1056/NEJMsa0911535
5. Patel J, Ahmed K, Guru KA, Khan F, Marsh H, Shamim Khan M, Dasgupta P. An overview of the use and implementation of checklists in surgical specialities - a systematic review. Int J Surg. 2014 Dec;12(12):1317-23. doi: 10.1016/j.ijsu.2014.10.031. Epub 2014 Oct 28. PMID: 25448652.
6. Madhu CR, Dourish P, Pratt W. Temporality in medical work: Time also matters. Comput Supported Coop Work. 2006 Feb;15(1):29–53. doi: 10.1007/s10606-005-9010-z

7. Fourcade A, Blache JL, Grenier C, Bourgain JL, Minvielle E. Barriers to staff adoption of a surgical safety checklist. BMJ Qual Saf 2012;21(3):191-7. doi: 10.1136/bmjqs-2011-000094
8. Hannam JA, Glass L, Kwon J, Windsor J, Stapelberg F, Callaghan K, Merry AF, Mitchell SJ. A prospective, observational study of the effects of implementation strategy on compliance with a surgical safety checklist. BMJ Qual Saf. 2013 Nov;22(11):940-7. doi: 10.1136/bmjqs-2012-001749. Epub 2013 Jul 9. PMID: 23840072.
9. Levy SM, Senter CE, Hawkins RB, Zhao JY, Doody K, Kao LS, Lally KP, Tsao K. Implementing a surgical checklist: more than checking a box. Surgery. 2012 Sep;152(3):331-6. doi: 10.1016/j.surg.2012.05.034. Epub 2012 Jul 6. PMID: 22770952.
10. Mahmood T, Haji F, Damignani R, Bagli D, Dubrowski A, Manzone J, Truong J, Martin R, Mylopoulos M. Compliance does not mean quality: an in-depth analysis of the safe surgery checklist at a tertiary care health facility. Am J Med Qual. 2015 Mar-Apr;30(2):191. doi: 10.1177/1062860614567830. Epub 2015 Jan 13. PMID: 25586102.
11. Kulp L, Sarcevic A, Cheng M, Zheng Y, Burd RS. Comparing the Effects of Paper and Digital Checklists on Team Performance in Time-Critical Work. Proc SIGCHI Conf Hum Factor Comput Syst. 2019 May;2019:547. doi: 10.1145/3290605.3300777. PMID: 31633126; PMCID: PMC6800573.
12. Kulp L, Sarcevic A, Farneth R, Ahmed O, Mai D, Marsic I, Burd RS. Exploring Design Opportunities for a Context-Adaptive Medical Checklist Through Technology Probe Approach. DIS (Des Interact Syst Conf). 2017 Jun;2017:57-68. doi: 10.1145/3064663.3064715. PMID: 30381804; PMCID: PMC6205694.
13. Sarcevic A, Zhang Z, Marsic I, Burd RS. Checklist as a Memory Externalization Tool during a Critical Care Process. AMIA Annu Symp Proc. 2017 Feb 10;2016:1080-1089. PMID: 28269905; PMCID: PMC5333210.
14. Fitzgerald M, Cameron P, Mackenzie C, Farrow N, Scicluna P, Gocentas R, Bystrzycki A, Lee G, O'Reilly G, Andrianopoulos N, Dziukas L, Cooper DJ, Silvers A, Mori A, Murray A, Smith S, Xiao Y, Stub D, McDermott FT, Rosenfeld JV. Trauma resuscitation errors and computer-assisted decision support. Arch Surg. 2011 Feb;146(2):218-25. PubMed PMID: 21339436
15. Huang AY, Joerger G, Fikfak V, Salmon R, Dunkin BJ, Bass BL, Garbey M. The SmartOR: a distributed sensor network to improve operating room efficiency. Surg Endosc. 2017 Sep;31(9):3590-3595. doi: 10.1007/s00464-016-5390-z. Epub 2017 Feb 24. PMID: 28236014.

Inclusion of Priority Populations in Research

Inclusion of Women and Minorities

The study population (trauma team leaders) was drawn from the staff working in trauma resuscitation at Children's National Medical Center. No gender or minority exclusions were used.

Inclusion of Children

The project collected data on patient and event attributes for 187 resuscitations that allowed us to evaluate the impact of checklist compliance on teamwork, process errors and patient outcomes.

Anticipated Recruitment

Study Title: Analysis of Factors Associated with Clinical Checklist Compliance

Total Planned Enrollment: 400

TARGETED/PLANNED ENROLLMENT: Number of Subjects			
Ethnic Category	**Females**	**Males**	**Total**
Hispanic or Latino	100	100	200
Not Hispanic or Latino	100	100	200
Ethnic Category: Total of All Subjects *	200	200	400
Racial Categories			
American Indian/Alaska Native	40	40	80
Asian	40	40	80
Native Hawaiian or Other Pacific Islander	40	40	80
Black or African American	40	40	80
White	40	40	80
Racial Categories: Total of All Subjects *	200	200	400

* The "Ethnic Category: Total of All Subjects" must be equal to the "Racial Categories: Total of All Subjects."

Actual Recruitment

Study Title: Analysis of Factors Associated with Clinical Checklist Compliance

Total Planned Enrollment: 187

TARGETED/PLANNED ENROLLMENT: Number of Subjects			
Ethnic Category	**Females**	**Males**	**Total**
Hispanic or Latino	46	47	93
Not Hispanic or Latino	47	47	94
Ethnic Category: Total of All Subjects *	93	94	187
Racial Categories			
American Indian/Alaska Native	7	10	17
Asian	20	20	40
Native Hawaiian or Other Pacific Islander	10	20	30
Black or African American	25	25	50
White	25	25	50
Racial Categories: Total of All Subjects *	87	100	187

* The "Ethnic Category: Total of All Subjects" must be equal to the "Racial Categories: Total of All Subjects."