

Checklist Design Reconsidered: Understanding Checklist Compliance and Timing of Interactions

Leah Kulp¹, Aleksandra Sarcevic¹, Yinan Zheng², Megan Cheng², Emily Alberto², Randall Burd²

Drexel University
Philadelphia, PA, USA
{ljk58,as3653}@drexel.edu

Children's National Medical Center
Washington DC, USA
{yzheng,mcheng,ealberto,rburd}@childrensnational.org

ABSTRACT

We examine the association between user interactions with a checklist and task performance in a time-critical medical setting. By comparing 98 logs from a digital checklist for trauma resuscitation with activity logs generated by video review, we identified three non-compliant checklist use behaviors: failure to check items for completed tasks, falsely checking items when tasks were not performed, and inaccurately checking items for incomplete tasks. Using video review, we found that user perceptions of task completion were often misaligned with clinical practices that guided activity coding, thereby contributing to non-compliant check-offs. Our analysis of associations between different contexts and the timing of check-offs showed longer delays when (1) checklist users were absent during patient arrival, (2) patients had penetrating injuries, and (3) resuscitations were assigned to the highest acuity. We discuss opportunities for reconsidering checklist designs to reduce non-compliant checklist use.

Author Keywords

Checklist design; medical checklist; dynamic checklist; mixed methods; video review; trauma resuscitation.

CSS Concepts

• **Human-centered computing~Human computer interaction (HCI); Empirical studies in HCI.**

INTRODUCTION

Following successful application of checklists in aviation, many clinical settings have adopted this cognitive aid with the aim of reducing errors and improving patient care [2,6,17,19,28,44,48]. Prior work on the effects of medical checklists, however, has shown mixed results, depending on clinical domain, checklist type, study design, participation rate, and user training. Multiple studies have found full compliance with checklists, leading to positive effects on patient outcomes through improved protocol adherence and time to task completion [11,26,30,38,46]. In contrast, other

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

research found that checklists negatively impact medical work, attributing those effects to low checklist compliance due to lack of user training or poor checklist design [12,45]. This prior work has highlighted a knowledge gap about particular components of work processes or checklist use that may be affecting outcomes, calling for a more nuanced approach to studying checklist compliance.

Checklist compliance has been defined as the number of completed sections or items [36], as a combination of completion and accuracy [41], and whether or not the checklist was actually used [14]. To understand compliance, most prior studies used pre- and post-implementation design, collecting data through surveys and retrospective analyses [14,18,27,31,43,46]. Checklist success measures have mostly focused on adherence to protocols, mortality rates, and infection or complication rates. While 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 affects checklist use.

In this work, we designed and deployed a digital checklist for pediatric trauma resuscitation—a dynamic medical process of rapidly identifying and treating life-threatening injuries in children—to better understand user interactions with a checklist and how the nature of tasks affected those interactions. The resuscitation room provided an ideal setting for studying the relationship between checklist compliance and task performance because the checklist is administered concurrently with clinical activities performed by an interdisciplinary team under time pressure. Our goal for this study was twofold. First, we determined the types of non-compliant checking behaviors and task features contributing to those behaviors. Second, we analyzed how the timing of checking items correlated to completion times of the corresponding tasks. We obtained data about user interactions from 98 checklist logs, collecting timestamps for all checked items and written notes about patient status. These digital traces of checklist use, along with the ground truth data (activity logs with start and finish times for all tasks) and qualitative observations from videos provided a unique dataset for understanding user interactions with the checklist and the contexts within which they occurred.

Our measure of compliance has been 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. Unlike prior studies of checklist compliance, our analyses uncovered associations between actual task performance, timing of check-offs, and different contexts in which the checklist was used, adding new insight to our current knowledge about checklist compliance in complex work settings.

RELATED WORK

Checklist use and design have been topics of research in many disciplines, including human factors, human-computer interaction, and health sciences. Below we review two areas of research relevant to our current work.

Studies of Checklist Use and Compliance

Prior studies of compliance with medical checklists have evaluated adherence to care protocols [1,16,32,39,46], task completion rates and task accuracy [14,27,31,43], and timeliness of task performance [18,20,22,23]. Mainthia et al. [32], for example, evaluated the efficacy of an electronic checklist system that reminds surgical teams to pause before proceeding with surgery, finding a significant increase in checklist compliance and adherence to procedures. In contrast, Hulfish et al. [20] studied the impact of a trauma checklist displayed on a wall monitor, showing improvements in completion rates for only three of 30 tasks, and no difference in timing of task performance. Although checklist use has often been associated with improved protocol adherence and task completion rates, compliance with checklists has been inconsistent. Prior research has shown that checking off items on a checklist does not always correspond with task performance—care providers check off items without completing corresponding tasks or perform a task without checking the associated checklist box [14,23,29,43,46]. A study evaluating task performance using medical record review found that only 54% of surgical tasks were completed despite 85% of checklist items being checked [43]. In another study, direct observation of checklist completion during surgical procedures found that corresponding tasks were performed between 0% to 97% per item, even when all checklist boxes were checked [27]. Kulp et al. [23] found that the introduction of an electronic checklist for trauma resuscitation did not impact task performance, but improved checklist completion rates.

Two aspects of checklist compliance that remain unexplored include the relationship between compliance behaviors and task performance, and how different task features affect checklist compliance. For example, we have limited knowledge about the types of tasks that are associated with missing or delayed check-offs. We also lack insights into how long it takes to check off items after task completion and what factors contribute to this timing. Our unique dataset (activity logs, time-stamped checklist interactions, and contextual data about cases) has allowed us to go beyond simple completion rates and study how the nature of tasks

shapes checklist compliance. This nuanced insight into compliance is critical because it has important implications for improving checklist designs to better facilitate dynamic work processes, especially those that depend on checklists as a safety procedure (e.g., airplane cockpits, air traffic control rooms, nuclear power station control rooms). Some work settings may also require that checklists are used concurrently with ongoing tasks, which introduces new complexity and potentially increases mental load and stress [47]. Non-compliant checklist use in these settings may provide false reassurance that tasks have been performed or skipped, posing widespread risks to safety.

Studies of Documentation and Timeliness

Many studies in HCI and health sciences have focused on completeness and timeliness of information documentation in clinical settings [7,9,10,13,21,33,34,37,40,41], showing mixed results. Østerlund et al. [33], for example, described how an electronic documentation system supported collaboration among care providers by allowing for spatial and temporal details of patient care in the record. Deering et al. [11] introduced a standard checklist of key elements that should be included for complex birth deliveries and added it to an EHR, which led to significant improvements in documentation. Other studies, however, found misalignments between EHR design and actual clinical practices [3,9,37]. Chen [9], for instance, described how the implementation of an EHR did not have a significant effect on clinical performance, forcing the use of transitional documents to fill the gaps between the EHR and actual patient care. Pine and Mazmanian [37] described perfect but inaccurate accounts of nurse documenters as the result of EHR implementation. A few studies have also focused on understanding how timeliness of documentation correlated with clinical task performance, finding that only 8% of tasks were documented in real time, while 12% of were documented in a delayed manner [21].

Prior work on checklist compliance and documentation in healthcare has found low checklist compliance, incomplete and inaccurate checklist use, delayed documentation of task performance, and misalignments between new technology and complex workflows. We contribute to this literature, and to HCI in particular, by (1) identifying task features that contribute to non-compliant checklist use, (2) determining the timing of checklist use in relation to task performance, and (3) proposing new approaches to checklist design to support complex and dynamic work processes.

METHODS

Our digital checklist has been in use during resuscitations at a regional pediatric, level 1 trauma center since October 2016. During the study period (January 1, 2017-March 1, 2018), the center treated 611 trauma patients. Of these, 197 cases had a signed consent and video files available for research purposes. We excluded cases that did not have ground truth activity data and checklist log data available, leaving us with 98 cases in the dataset. We found no

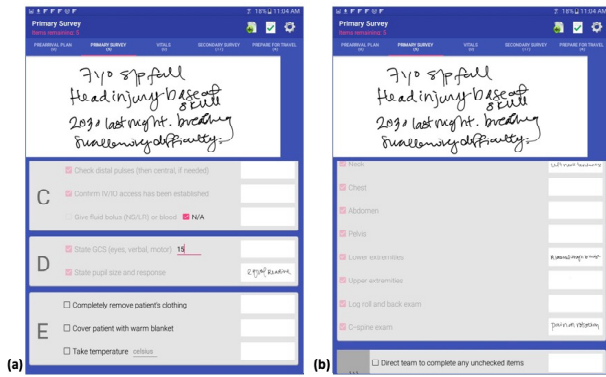


Figure 1: Example screens from the digital checklist with user notes in the margin area, typed and stylus notes on the primary survey (a), and checks/notes on the secondary survey (b).

significant differences between the clinical features of the 98 selected cases and those of 197 available cases, supporting that the selection of cases was not biased. The study was approved by the Legal and Risk Management Department and the Institutional Review Board at the hospital.

Research Setting

Upon receiving a pager notification of the pending arrival of an injured patient, a multidisciplinary team of physicians and nurses assembles to rapidly evaluate and stabilize the patient. At our research site, trauma teams are activated into one of three levels based on the severity of patient injury: “stat” (standard level acuity), “attending” (highest level acuity), or “transfer” (patient arriving from another institution). Before patient arrival, the team prepares the room based on pre-hospital information. After arrival, the patient is evaluated and treated following a sequence of tasks based on the ATLS protocol [4]. At the end of the resuscitation, the team develops a plan for definitive care. The ATLS protocol has two parts. The primary survey evaluates and manages major physiological systems, including airway evaluation, breathing assessment, circulatory status assessment, neurological or “disability” assessment, and exposure of patient (ABCDE). The secondary survey identifies other injuries using a head-to-toe physical exam. The core team includes five to seven care providers but can expand to up to 12 members, depending on the type and severity of the patient’s injuries. One of the team members is responsible for leading the team and delegating tasks to ensure the team’s adherence to protocols and treatment plans. This role is assigned to a senior surgery resident, a surgery fellow, or an attending surgeon. Other team members include a junior surgery resident or nurse practitioner, who are responsible for evaluating the patient and performing the physical evaluation, as well as bedside nurses who assist with the administration of medications and other treatments.

To support the leader in their role, the surgical leadership at our research site introduced a paper checklist for trauma resuscitation in 2012 [35]. This 50+ item checklist contains four major sections, each corresponding to a phase of care:

pre-arrival plan, primary survey and vital signs, secondary survey, and departure plan. When leaders arrive at the trauma bay, they start using the checklist as the team prepares for patient arrival and later performs exams and treatments. This concurrent checklist administration is usually 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.

Digital Checklist and Interaction Logs. Our team designed and implemented a digital version of the paper checklist in October 2016 [23,24]. The leaders can choose between a paper or digital checklist format, depending on their preferences. As of December 2019, the digital checklist has been used in more than 550 trauma resuscitations. This checklist application mirrors the design of its paper counterpart, with several feature enhancements afforded by tablet computers. The application includes the same sections, separated into five tabs (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.

Data Collection

We used a multi-step process to collect and prepare data for comparing the checklist logs and ground truth activity logs. We first transcribed the checklist logs and coded videos to create activity logs. Prior to comparing the log files, we matched the checklist items to corresponding tasks on the activity logs. We also collected context information from a clinical database to assess associations between different contexts and checklist use.

Ground Truth Coding and Activity Logs. To define resuscitation tasks and determine performance attributes, the domain experts on our team first created a data dictionary for about 260 resuscitation tasks and activities. Medical experts at the research site then used a video annotation software to review videos from three camera angles and to code and timestamp task performance. Using the data dictionary, the coders reviewed videos of 98 cases, tracking each team member throughout the resuscitation and coding each resuscitation task as it was performed. 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, as described next.

Matching Checklist Items with Task Performance. For this study, we selected 32 (out of 55) checklist items that correspond to clinically relevant, time-critical pre-arrival, assessment and treatment tasks: three in the pre-arrival plan section, 11 in the primary survey, four vital sign items, and 14 items in the secondary survey (Table 2). To understand how checking these items correlated with the actual task performance, we first determined which tasks from the activity log best matched each checklist item. For some items, matching was possible based on one-to-one association. For example, for the pre-arrival plan item “Oxygen connected to NRB [non-rebreather mask],” the corresponding task from the activity log is “Oxygen preparation” and the check-off timing was compared to the start and finish times recorded in the activity logs for this task. For other checklist items, matching was more complex because several activities may indicate task performance. For instance, the primary survey checklist item “State GCS” requires the team to assess the patient’s neurological status by evaluating the patient’s eye, verbal, and motor responses using a three-part score of the Glasgow Coma Scale (GCS). Several tasks may signal the start of the GCS activity, including eye assessment, verbal assessment, motor assessment, as well as verbalization of the eye, verbal or motor scores. To consider this activity performed to completion, a physician must verbalize either the total GCS score (all three values added together) or all three values individually. The matching process was done iteratively and in collaboration with medical experts at the hospital.

Context Information. From the hospital’s trauma registry, we selected six resuscitation features for assessment of context effects on checklist completion: team activation level (stat, attending, transfer), team leader experience level (fellow or senior resident), mechanism of injury (blunt, penetrating, burn, or other), whether the team was notified ahead of the patient’s arrival or not (early notification vs. “now” activation), whether the team leader was present before the patient arrived, and time of day (day/night). Because these six features are related to checklist users (leaders) and team performance, they are more likely to affect checklist use.

Data Analysis

Data analysis proceeded in four phases: (1) analyzing item check-offs and task performance data to determine checklist compliance, (2) comparing timestamps between item check-offs and task completion to determine the timing of check-offs, (3) statistical analysis to identify associations between different contexts and non-compliant checklist use, and (4) video analysis to further unpack non-compliant checklist use.

Checklist Compliance Analysis. 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 also performed a univariate analysis to determine which contexts were associated with non-compliant checklist use (i.e., falsely checked items for tasks that were not performed or were started but not completed, and failed checks for tasks that were performed to completion).

Timestamp Extraction and Comparison. 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 time points when check-offs occurred in relation to task performance: before, during, or after task performance.

Timing of Check-Offs and Context. To determine associations between the timing of check-offs and task performance, we first calculated task duration (the time needed to perform a task from start to finish) for each of the 32 tasks in 98 cases (Table 4, Task Duration column). We then compared the start times for each task performed to completion to timestamps for corresponding item check-offs that occurred during or after task completion. These comparisons showed the time difference between task performance and item check-offs (Table 4, Check-off Delay column). Because the data were right skewed, we used negative binomial regression to determine associations between the six contexts and timing of checkoffs.

Video Review. To further understand factors contributing to non-compliant checklist use, we reviewed videos of the 98 cases. In each video, we observed the leaders’ behaviors and team communications surrounding the tasks that had high frequency of non-compliant check-offs. A physician at the research site provided clinical guidance. We then analyzed the observations from the ground up to identify behavior patterns and factors that affected those behaviors.

RESULTS

We first review contextual information about the cases. We then report on the five checking behaviors (two compliant and three non-compliant) and associations with different contexts. Finally, we analyze the timing of check-offs and how different tasks and contexts contributed to this timing.

General Case Observations

Our dataset includes 98 video-reviewed and coded trauma patients, with a mean age of 6.5 (SD = 4.9). Most patients were male (66%). The majority of patients (94%) were injured by a blunt mechanism, two patients (2%) by a penetrating mechanism, and four (4%) by other mechanisms. Fifty-nine patients were triaged as a “stat” activation (60%), seven patients as an “attending” activation (7%), and 32 patients as “transfers” (33%). Most patients arrived after the trauma team had been notified (91%). Only nine patients arrived at the emergency department without prior team notification, i.e., “now” activations (9%). Fifty-two percent of resuscitations were led by a surgical fellow, while the remaining cases had a senior resident in the team leader role (48%). Daytime (57%) and weekday trauma activations (66%) were more common than those occurring after hours.

Checklist Compliance Behaviors

We identified five checklist compliance outcomes (Table 1). *True checks* (item was checked and corresponding task was performed to completion) and *true non-checks* (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 non-compliant behaviors included *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 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 non-checks and 49% were failed checks. We next describe each of the checking behaviors in greater detail.

Compliant Behaviors—True Checks and Non-Checks. The true checks accounted for 79% of all checked items. Seven of the top ten items that were correctly checked when their corresponding tasks were completed came from the secondary survey. For example, head exam was completed in 92 cases and checked in 91, pelvis exam was completed in 83 cases and checked in 82, and neck exam was completed in 70 cases and checked in 69. The remaining three items were in the primary survey (e.g., pulses were assessed in 58 and checked off each time, blood pressure was assessed in 98 cases and checked in 96, and GCS was assessed and stated in 90 cases and checked in 88). Because not all items on the checklist need to be performed for all patients, we also found that leaders accurately documented non-applicable or not performed tasks by not checking their corresponding items four percent of the time (140 true non-checks of possible 3,136 checks). The most frequent true non-checks were found in the pre-arrival plan and primary survey sections. For example, not all patients require preparatory tasks like hooking up the suction equipment and placing warming

Table 1: Five checklist compliance outcomes based on checklist use and task performance.

	Item checked	Item not checked
Task performed to completion	True check	Failure to check
Task not performed	False check	True non-check
Task started but not performed to completion	Inaccurate check	

equipment (Bair Hugger™) on the bed. Similarly, fluid bolus or blood are only administered in high acuity cases, which explains our finding that the “Give fluid bolus or blood” task was not performed and thus left unchecked in 88 cases.

Non-Compliant Behaviors—False Checks. False checks 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). For example, the “Ocular/periorbital integrity” item had 94 false checks, indicating that this assessment of patient eyes and surrounding areas was not performed in 94 cases. The item “Confirm IV/IO access has been established” was falsely checked 48 times, implying that nurses did not establish intravenous access for 48 patients. After reviewing videos for team- and leader behaviors, we identified several patterns that explain these false, non-compliant 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. Similarly, nose and mouth are often assessed as part of the overall face exam (e.g., the bedside physician would report “Facial bones are stable, no blood in nasal- or oropharynx”), even though the clinical practice guidelines recommend separating these tasks. Accordingly, the physician leaders checked off these items in more than 30 cases, while the coders marked those tasks as not performed. 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. For example, clinical practice guidelines require fully covering patients (above the nipples) with a warm blanket. The leaders, however, falsely checked

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

Task	False Checks (n=452)	Inaccurate Checks (n=154)	Failed Checks (n=134)
<i>Pre-arrival Plan</i>			
Oxygen Equip.	6	0	14
Suction Equip.	8	0	11
Bair Hugger	6	1	9
<i>Primary Survey</i>			
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
<i>Secondary Survey</i>			
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

this item in 19 cases when the blanket was placed on the bed but it did not sufficiently cover the patient. Or, in cases where establishing IV access was checked off but the task was coded as not performed, we observed that these patients had IV access already established upon their arrival.

The univariate analysis of associations between the six contexts and false checks showed that leaders were more likely to false-check the checklist items when they were present at the time of patient arrival ($p=0.003$) and when patients sustained a blunt injury ($p=0.0089$) (Table 3).

Non-Compliant Behaviors—Inaccurate Checks. The inaccurate checks accounted for 5% of all checked items, occurring for 14 of the 32 tasks (Table 2). The items with most inaccurate check-offs included “Check distal pulses” item on the primary survey (38 inaccurate checks) and then “Mouth” (32), “Nose” (18), and “Upper extremities” (16) on the secondary survey. Frequent inaccurate checks were also found for the primary survey items “Confirm airway is protected” (14) and “Completely remove patient’s clothing” (10). Our video review showed that inaccurate checks mostly

occurred when team leaders checked off items for tasks that were started but not completed. For the pulse assessment task, for instance, 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. Although pulses were only partially assessed in these cases, the leaders still checked them off. The same factors contribute to inaccurate check-offs for the upper extremities exam on the secondary survey. Physicians should assess both sides, but they only fully evaluate one side because other team members occupy the other side of the body. Team leaders also inaccurately checked off Mouth, Nose, Ears, and Neck items on the secondary survey before these tasks were performed to completion. Our analysis of five inaccurate checks for the “Confirm IV/IO access has been established” showed a different pattern: because this is a multi-step activity that requires several steps and sometimes multiple attempts to establish the IV line, team leaders prematurely checked off the item when they observed nurses starting this task, rather than waiting for confirmation that the task was completed.

Non-Compliant Behaviors—Failed Checks. 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). The most frequent items with failed checks included “Take temperature” (15 failed checks) and “Give fluid bolus or blood” (10) on the primary survey, and two preparatory tasks in the pre-arrival plan section (“Oxygen connected to NRB” (14) and “Suction hooked up” (11)) (Table 2). For the “Take temperature” task, we observed through video that team leaders often missed the verbalizations of the temperature value due to noise. The “Give fluid bolus or blood” item is among the few checklist items that do not apply to all patients. To minimize the number of user clicks, the “N/A” boxes associated with these items are checked off by default (Figure 1(a)); unchecking the “N/A” boxes when the task is applicable automatically checks off the item. Through video review, we found that that leaders in the 10 cases when fluid and/or blood were indeed administered did not uncheck the “N/A” box. For the pre-arrival items, we observed they were left unchecked when the leader arrived after the patient. In these cases, the leader began using the digital checklist at the point where the team was and did not retroactively check off the pre-arrival items.

The univariate analysis of associations between the six contexts and failed checks showed that leaders were more likely to miss checking an item for completed tasks when they were absent at the time of patient arrival ($p<.001$), when patients had non-blunt injuries (penetrating or other) ($p<.001$), and when activations occurred during the 7am-7pm, daytime shift ($p=0.02$) (Table 3).

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

Table 3: Results from univariate analysis showing association between context elements and false checks for incomplete tasks and failed checks for completed tasks.

Context Elements	False Checks				Failed Checks			
	True non-checks (n = 140)	False checks (n = 606)	Among incomplete (n = 746)	p-value	True Check (n = 2256)	Failure to check (n = 134)	Among complete (n = 2390)	p-value
Team leader experience % (fellow, senior resident)	55	45.9	47.6	0.051	53.5	52.2	53.4	0.78
Injury mechanism % (blunt, not blunt)	90.7	96	95	0.0089	94.6	75.4	93.5	<.0001
TL present at pt. arrival % (present, not present)	70	81.2	79.1	0.003	76.7	58.96	75.7	<.0001
Time of day % (daytime, nighttime)	55.7	55.5	55.5	0.95	57.1	67.2	57.7	0.022
“Now” activation % (now, not now)	13.6	8.8	9.7	0.081	8.95	10.5	9	0.56
Activation level %								
<i>Transfer</i>	37	40.4	39.8	0.56	30.8	23.9	30.4	0.14
<i>Attending</i>	5	6	6.2		7.3	10.5	7.5	
<i>Stat</i>	57.9	53	54		61.9	65.7	62.1	

items in relation to task performance but also *what factors contributed to the variable timing*. The digital checklist and available logs allowed us to perform this temporal analysis with precision for all true checks (2,256 out of 2,862 items).

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” (Figure 2(a)). Our video review of leader- and team behaviors showed that many of these tasks are multi-step and take longer to perform. For example, establishing IV/IO access requires that a nurse first prepares the patient’s arm for IV placement, establish IV access (a step that can take multiple attempts) and then confirm the IV placement was successful. The back exam similarly includes several steps, starting with a verbal count initiating the log roll, log rolling the patient onto their side, and then performing the exam. We found that team leaders checked off these tasks as soon as they ordered them, anticipating they would be performed to completion.

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,” “State GCS,” “Ears” exam, “Lower Extremities” exam, and “Confirm airway is protected” (Figure 2(b)). Similar to pre-checked tasks, tasks checked off during performance are often longer in duration

and may be performed more than once. For example, c-spine stabilization is one of the longest tasks because it is performed throughout the entire resuscitation and can be a combination of applying a cervical collar or manually stabilizing the patient’s c-spine (Table 4).

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. Vital signs assessments and pre-arrival plan tasks were among the most frequently post-checked items (Figure 2(c)). Patient vital signs show up on a monitor once the monitor is attached to the patient, making these tasks the shortest in duration (one second only, Table 4). Team leaders, however, do not always check off these items immediately after they show up on the monitor. The preparatory tasks are also sometimes checked after they are completed due to the leaders’ late arrival.

While post-task checking is typical for checklists, it was important to also determine the point at which post-checks became delayed, non-compliant behaviors, i.e., whether items were checked within a reasonable time period (slightly after task completion) or checked long after task completion. To answer this question, we 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) (Table 4). For each task, we excluded pre-checks (non-compliant 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

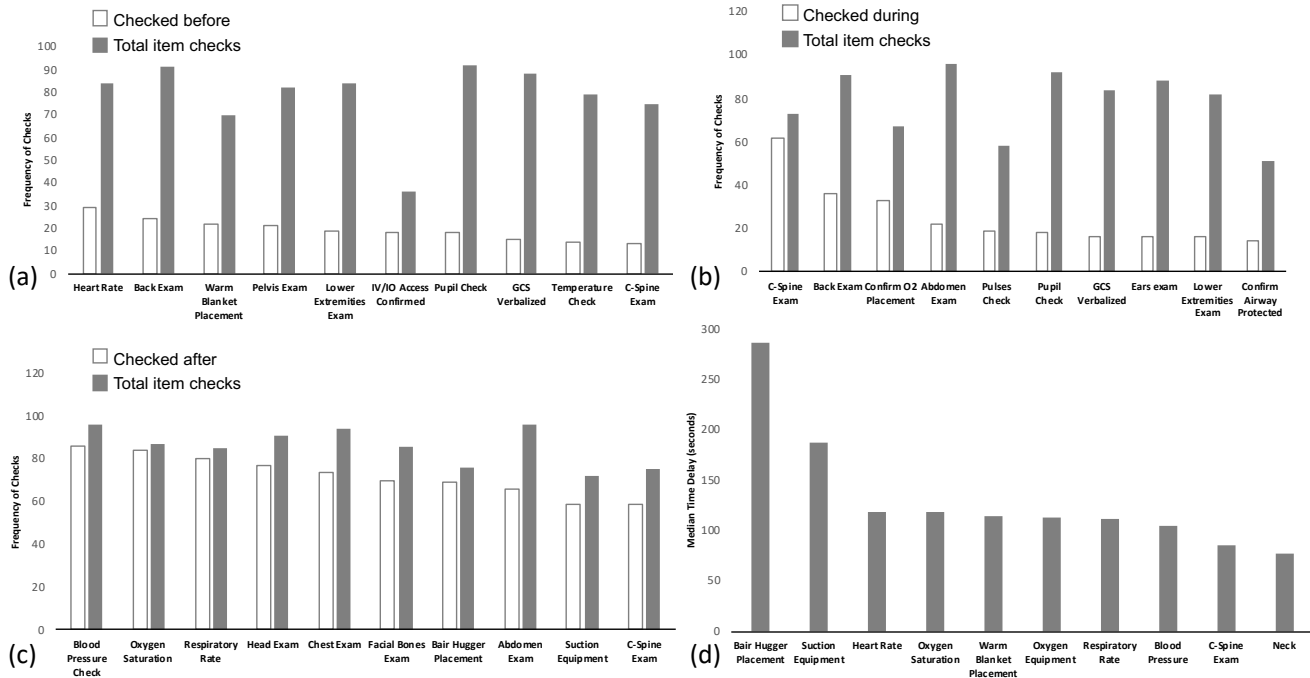


Figure 2: Timing of item check-offs categorized as (a) checked before task start, (b) during task performance, and (c) after task completion. Items that were checked with longest delay after task performance are shown in (d).

C-spine and Neck exams on the secondary survey (Table 4, Figure 2(d)). We considered these check-offs delayed because it took the leaders close to 2 minutes or longer to check off the items. Although a 2-minute delay may not have an impact on a slow-moving process with a single task performer, it could lead to negative outcomes for concurrent checklist administration in safety-critical work settings.

Associations between Delayed Check-Offs and Contexts

The regression analysis showed that leaders' late arrival was associated with longer delays for the overall resuscitation, as well as for primary and secondary surveys ($p=0.0016$, $p<.001$, $p=0.01$, respectively). Overall and in the secondary survey, patients with blunt injuries were associated with shorter check-off delays than patients with penetrating and other injuries ($p=0.03$, $p=0.01$, respectively). In the secondary survey, attending level activations had significantly longer check-off delays compared to stat level activations ($p=0.03$). No significant differences were found between checkoff delays among senior fellows and residents, among daytime and nighttime cases, or among "now" activations and activations with pre-notifications.

DISCUSSION

We compared interaction timestamps from digital checklist logs to task start and finish times from ground truth activity logs to better understand the association between user interactions with a checklist and task performance in a dynamic medical setting. We found that checklist items were falsely checked 16% of the time and inaccurately checked 5% of the time. Checklist users also failed to check off items

when their corresponding tasks were performed 7% of the time. Our analyses have shown that the three non-compliant checklist behaviors were not random, but rather 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. Similar misalignments have already been found and discussed in HCI and CSCW literature. As mentioned earlier, Pine and Mazmanian [37] described how the inflexible nature of medical systems created a conflict between documenting what was actually done and documenting what the system required (e.g., establishing an IV line before receiving approval from the pharmacy to insert an IV). We observed a similar conflict in checklist use during trauma resuscitations. 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. Additionally, changes in patient status and emerging contextual information often force teams to deviate from the standard order of tasks on the checklist. While some variability is common, prior research has identified factors associated with

Table 4: Median task durations and check-off delays.

Activities	N	Task Duration Median (Q1, Q3)	Check-off Delay Median (Q1, Q3)
<i>Pre-arrival Plan</i>			
Oxygen Equip.	70	51s (37, 81)	113s (29, 220)
Suction Equip.	63	23s (12, 39)	187s (70, 289)
Bair Hugger	69	3s (2, 15)	286s (184, 367)
<i>Primary Survey</i>			
Airway assessment	72	6s (4, 11)	31s (2, 91)
C-spine stabilization	64	1057s (598, 1478)	-964s (-1366, -466)
Confirm O ₂	58	89s (51, 152)	-27s (-102, 49)
IV/IO Access	18	208s (127, 343)	-6s (-108, 160)
Pulses	53	29s (18, 44)	12s (-5, 71)
GCS verbalized	73	13s (3, 22)	13s (1, 59)
Pupils	74	7s (4, 11)	13s (0, 76)
Remove Clothing	39	52s (19, 123)	38s (-9, 123)
Warm Blanket	48	9s (7, 12)	114s (65, 248)
Temperature	65	16s (14, 28)	48s (4, 115)
Heart Rate	55	1s (1, 1)	119s (53, 202)
Respiratory Rate	80	1s (1, 1)	112s (58, 161)
Oxygen Saturation	84	1s (1, 1)	118s (74, 196)
Blood Pressure	94	27s (20, 37)	105s (43, 185)
<i>Secondary Survey</i>			
Head	83	8s (5, 15)	64s (20, 117)
Ears	74	22s (14, 30)	31s (1, 119)
Ocular Integrity	---	---	---
Facial Bones	77	7s (4, 12)	59s (18, 117)
Nose	32	1.7s (1, 3)	57s (21, 173)
Mouth	26	4s (3, 6)	78s (22, 155)
Neck	59	4s (2, 5)	48s (20, 98)
Chest	83	5s (3, 8)	43s (9, 98)
Abdomen	88	6s (3, 11)	35s (0, 83)
Pelvis	61	3s (2, 4)	48s (7, 91)
Lower extremities	65	15s (9, 28)	28s (1, 83)
Upper extremities	58	17s (9, 32)	42s (2, 102)
Back exam	67	34s (25, 54)	-3s (-22, 47)
C-spine exam	62	6s (4, 11)	86s (20, 163)

harmful variability in the resuscitation process, including delayed and omitted tasks (e.g., [42]). Although the checklist for trauma resuscitation was introduced precisely to address this harmful variability, completion rates remain suboptimal and other non-compliant behaviors persist.

The task length and complexity also affected how users interacted with the checklist. Although some prior studies of checklist compliance looked at the relationship between compliance and task performance [14,23,43], the nature of tasks and contexts in which those tasks were performed were rarely examined. Sparks et al. [43], for example, evaluated both compliance with surgical safety checklists (whether boxes were checked) and accuracy of the associated tasks (whether tasks were performed to completion), but did not consider task features and their effects on checklist interactions. 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., multi-step). Although our non-compliant 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. The question then is how can the HCI community address this challenge and reduce non-compliant checklist behaviors?

Checklist Design Reconsidered

Checklists in aviation and healthcare have been designed to match user needs in specific settings [8], where they provide a step-by-step approach to addressing the issues [49]. We challenge this prior work on checklist design, suggesting a shift in design paradigms as checklists are evolving and transitioning from paper to electronic formats [5]. A recent HCI study has already considered a different approach to checklist design, using activity theory and focusing on decisional criteria instead of task-based processes [25]. As our study has shown, the dynamic nature of the resuscitation environment and work of trauma teams requires a checklist design that can adapt and support the actual work process. Our digital checklist was designed by directly translating protocol tasks onto the checklist, but this design approach did not succeed in supporting dynamic work practices. Rather, we propose an approach to checklist design that focuses on types of tasks (e.g., simple yes/no items, longer time tasks, multi-step tasks), where the design will more accurately reflect the work “as is.” We next discuss three checklist design directions to facilitate dynamic work: (1) supporting different types of tasks, (2) supporting retrospective checking, and (3) supporting timely documentation.

Supporting Different Types of Task Performance

Our study showed that non-compliant behaviors occurred during tasks that were performed as part of another, already completed task or grouped into larger tasks, tasks that had long durations, tasks that were performed more than once, tasks that had multiple steps, and tasks that were rarely performed. Because these task features are characteristic of many work settings and processes, we discuss implications for reconsidering checklist design in general, and not just for medical care. These design suggestions are intended to reduce the overall cognitive demand and negative redundancy by minimizing clutter and user steps, and by supporting the actual practice based on task types.

Tasks that are performed as a group. Prior work states that task descriptions on the checklist are critical for successful use because the item label must be concise and to the point, while also including seemingly redundant tasks, even if they are second nature to the checklist administrator [12,15]. We challenge this recommendation because we found that redundant checklist items often lead to non-compliant behaviors. We propose that items could be redundant when exams yield positive findings, so they allow for comprehensive response. For example, we found that the “Ocular integrity” exam on the secondary survey was truly performed in only one out of 98 cases because it was usually covered during other tasks, either during the pupils assessment in the primary survey or during a general face exam in the secondary survey. The current checklist design may therefore be too granular and separating items that are

usually performed as a group created negative redundancy. An alternative approach to designing to support tasks that are performed as a group is to include the main task only but show related tasks as sub-items. In cases where there is a positive exam finding on a sub-item, the team leader may want to write a note (e.g., laceration or missing teeth), so the checklist design should still support this action.

Tasks with long durations. Currently, the checklist design only supports checking and unchecking items, as well as note taking. After taking a closer look at the nature of tasks, we found that items were also checked during task performance, regardless of whether the task was performed to completion. For example, the c-spine stabilization task was often checked when the patient's c-spine was stabilized, but this is an ongoing task and is usually performed on and off throughout the entire resuscitation. To accurately reflect the actual task performance, these types of tasks could have an "on" and "off" switch on the checklist.

Tasks that are performed more than once. Some tasks are repeated throughout the resuscitation, such as vital sign assessments. The checklist design, however, is mostly static, precluding users to check off an item multiple times and later visualize the items that have been checked off more than once. Checklist design should therefore allow users to enter multiple values and check-offs, as well as indicate when was the last time a task was performed via a timestamp or label.

Tasks with multiple steps. We found that some non-compliant behaviors occurred for tasks that have multiple steps. For example, pulses exam was often not completed because physicians only evaluated pulses on one side of the patient body. For these items, the checklist should allow for recording the progress of tasks instead of forcing users to indicate that the task was completed.

Tasks that are rarely performed. Current checklist design has a few features for rarely occurring tasks. For example, tasks that are only performed for patients that require intubation are hidden in an expandable section with a label "If Intubating," and users can expand this section to view the items. The other design feature is an "N/A" box that is checked by default but should be unchecked if a task is performed. As we observed in this study, this feature created problems in the cases where this item needed a check off. Rather, the checklist could dynamically update this item based on values that are entered for other tasks.

Supporting Retrospective Checking

We observed two patterns of retrospective checking that the current checklist doesn't support: when team leaders check off items for tasks that occurred *before the patient arrived* or *before the leader arrived* at the trauma room. Because these retrospective checks inaccurately document the time of task performance, it is important to reconcile the tasks that are performed at different stages of care, e.g., en route to the hospital or before the leader arrives. The checklist design should therefore support data entry for the entire process,

regardless of where or when it started. Two approaches are possible from a design perspective: (1) providing a section for tasks performed before, during or after the main event, or (2) indicating whether a checklist item refers to a task that was performed before or after the main event.

Supporting Timely Documentation

Approaches to checklist administration have been described as "do-list" or "challenge-response" [12,35]. Our study showed that delayed check-offs were skewed towards the right (i.e., delays were occurring closer to task completion), indicating that leaders mostly used the checklist in a do-list manner—waiting for task completion or verbalizations, and then checking off the corresponding items. However, some items were checked off beyond just a few seconds after task completion. Some leaders chose to check off items at the end of the event based on their recall. This approach, however, poses a safety risk because it does not accurately represent the work process. We found that vital signs and pre-arrival tasks were checked off almost two minutes and longer after they were performed. Because resuscitation tasks need to be completed in an efficient and timely manner, the checklist can dynamically support this type of work by using the values previously entered by users. A visible timer showing the elapsed time can also be started when the leader checks off the first item on the list. These features could similarly apply to other time-critical processes.

CONCLUSION

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 work, we studied the extent to which user interactions with a digital checklist represent actual activity in an emergency medical setting. Our unique dataset with time-stamped checklist interactions and activity logs allowed us to go beyond simple completion rates and understand how the nature of tasks shaped checklist compliance. We identified a set of non-compliant checklist behaviors and unpacked the factors contributing to non-compliant checklist use. To assist complex work processes, new checklist designs should consider task attributes and contexts within which those tasks are performed. Our study has three limitations: (1) only 32 (of 55) checklist items were selected for analysis because of our focus on time-critical task performance; (2) only six context elements (of 20) were used; and (3) this is a single-site study and checklist use could differ at other centers. Our future work will expand the analyses to remaining checklist items and other context elements, as well as evaluate the proposed design suggestions for their effects on user interactions.

ACKNOWLEDGMENTS

This research has been funded under grant number 1R03HS026057-01 from the Agency for Healthcare Research and Quality (AHRQ), and has been partially supported by the National Library of Medicine of the National Institutes of Health under grant number 2R01LM011834-05 and the National Science Foundation under grant number IIS-1763509.

REFERENCES

- [1] Aalok V. Agarwala, Paul G. Firth, Meredith A. Albrecht, Lisa Warren, and Guido Musch. 2015. An electronic checklist improves transfer and retention of critical information at intraoperative handoff of care. *Anesth. Analg.* 120, 1 (January 2015), 96–104. DOI:https://doi.org/10.1213/ANE.0000000000000506
- [2] Alexander F. Arriaga, Angela M. Bader, Judith M. Wong, Stuart R. Lipsitz, William R. Berry, John E. Ziewacz, David L. Hepner, Daniel J. Boorman, Charles N. Pozner, Douglas S. Smink, and Atul A. Gawande. 2013. Simulation-based trial of surgical-crisis checklists. *New England Journal of Medicine* 368, 3 (January 2013), 246–253. DOI:https://doi.org/10.1056/NEJMs1204720
- [3] Joan S. Ash, Marc Berg, and Enrico Coiera. 2004. Some unintended consequences of information technology in health care: The nature of patient care information system-related errors. *J Am Med Inform Assoc* 11, 2 (2004), 104–112. DOI:https://doi.org/10.1197/jamia.M1471
- [4] ATLS Subcommittee, American College of Surgeons' Committee on Trauma, and International ATLS working group. 2013. Advanced trauma life support (ATLS®): the ninth edition. *J Trauma Acute Care Surg* 74, 5 (May 2013), 1363–1366. DOI:https://doi.org/10.1097/TA.0b013e31828b82f5
- [5] George S. Avrunin, Stefan C. Christov, Lori A. Clarke, Heather M. Conboy, Leon J. Osterweil, and Marco A. Zenati. 2018. Process driven guidance for complex surgical procedures. *AMIA Annu Symp Proc* 2018, (2018), 175–184.
- [6] Vicente Behrens, Roman Dudaryk, Nicholas Nedeff, Joshua M. Tobin, and Albert J. Varon. 2016. The Ryder cognitive aid checklist for trauma anesthesia. *Anesth. Analg.* 122, 5 (May 2016), 1484–1487. DOI:https://doi.org/10.1213/ANE.0000000000001186
- [7] Marc Berg. 1999. Accumulating and coordinating: Occasions for information technologies in medical work. *Computer Supported Cooperative Work (CSCW)* 8, 4 (December 1999), 373–401. DOI:https://doi.org/10.1023/A:1008757115404
- [8] Barbara K. Burian, Anna Clebone, Key Dismukes, and Keith J. Ruskin. 2017. More than a tick box: Medical checklist development, design, and use. *Anesth. Analg.* (July 2017). DOI:https://doi.org/10.1213/ANE.0000000000002286
- [9] Yunan Chen. 2010. Documenting transitional information in EMR. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*, ACM, New York, NY, USA, 1787–1796. DOI:https://doi.org/10.1145/1753326.1753594
- [10] Lars Rune Christensen. 2016. On intertext in chemotherapy: an ethnography of text in medical practice. *Comput Supported Coop Work* 25, 1 (February 2016), 1–38. DOI:https://doi.org/10.1007/s10606-015-9238-1
- [11] Shad H. Deering, Kyle Tobler, and Rebecca Cypher. 2010. Improvement in documentation using an electronic checklist for shoulder dystocia deliveries. *Obstet Gynecol* 116, 1 (July 2010), 63–66. DOI:https://doi.org/10.1097/AOG.0b013e3181e42220
- [12] Asaf Degani and Earl L. Wiener. 1993. Cockpit checklists: Concepts, design, and use. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 35, 2 (June 1993), 345–359. DOI:https://doi.org/10.1177/001872089303500209
- [13] Gunnar Ellingsen and Eric Monteiro. 2003. Mechanisms for producing a working knowledge: Enacting, orchestrating and organizing. *Information and Organization* 13, 3 (July 2003), 203–229. DOI:https://doi.org/10.1016/S1471-7727(03)00011-3
- [14] Aude Fourcade, Jean-Louis Blache, Catherine Grenier, Jean-Louis Bourgain, and Etienne Minvielle. 2012. Barriers to staff adoption of a surgical safety checklist. *BMJ Qual Saf* 21, 3 (March 2012), 191–197. DOI:https://doi.org/10.1136/bmjqs-2011-000094
- [15] Eliot Grigg. 2015. Smarter clinical checklists: How to minimize checklist fatigue and maximize clinician performance. *Anesth. Analg.* 121, 2 (August 2015), 570–573. DOI:https://doi.org/10.1213/ANE.0000000000000352
- [16] Tobias Grundgeiger, M. Albert, Daniel Reinhardt, Oliver Happel, Andreas Steinisch, and Thomas Wurmb. 2016. Real-time tablet-based resuscitation documentation by the team leader: evaluating documentation quality and clinical performance. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 24, (2016), 51. DOI:https://doi.org/10.1186/s13049-016-0242-3
- [17] Brigitte M. Hales and Peter J. Pronovost. 2006. The checklist--a tool for error management and performance improvement. *J Crit Care* 21, 3 (September 2006), 231–235. DOI:https://doi.org/10.1016/j.jcrc.2006.06.002
- [18] Jacqueline Amy Hannam, L. Glass, J. Kwon, John Windsor, F. Stapelberg, Kathleen Callaghan, Alan F. Merry, and Simon J. Mitchell. 2013. A prospective, observational study of the effects of implementation strategy on compliance with a surgical safety checklist. *BMJ Qual Saf* 22, 11 (November 2013), 940–947. DOI:https://doi.org/10.1136/bmjqs-2012-001749
- [19] Elaine M. Hart and Harry Owen. 2005. Errors and omissions in anesthesia: A pilot study using a pilot's checklist. *Anesth. Analg.* 101, 1 (July 2005), 246–250. DOI:https://doi.org/10.1213/01.ANE.0000156567.24800.0B

- [20] Erin Hulfish, Maria Carmen G. Diaz, Megan Feick, Catherine Messina, and Glenn Stryjewski. 2018. The impact of a displayed checklist on simulated pediatric trauma resuscitations. *Pediatr Emerg Care* (February 2018). DOI:https://doi.org/10.1097/PEC.0000000000001439
- [21] Swathi Jagannath, Aleksandra Sarcevic, Victoria Young, and Sage Myers. 2019. Temporal rhythms and patterns of electronic documentation in time-critical medical work. ACM, 334. DOI:https://doi.org/10.1145/3290605.3300564
- [22] Deirdre C. Kelleher, Elizabeth A. Carter, Lauren J. Waterhouse, Samantha E. Parsons, Jennifer L. Fritzeen, and Randall S. Burd. 2014. Effect of a checklist on advanced trauma life support task performance during pediatric trauma resuscitation. *Acad Emerg Med* 21, 10 (October 2014), 1129–1134. DOI:https://doi.org/10.1111/acem.12487
- [23] Leah Kulp, Aleksandra Sarcevic, Megan Cheng, Yinan Zheng, and Randall S. Burd. 2019. Comparing the effects of paper and digital checklists on team performance in time-critical work. In *2019 CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019)*, ACM, New York, NY, USA, Glasgow, Scotland, UK, 13. DOI:https://doi.org/10.1145/3290605.3300777
- [24] Leah Kulp, Aleksandra Sarcevic, Richard Farneth, Omar Ahmed, Dung Mai, Ivan Marsic, and Randall S. Burd. 2017. Exploring design opportunities for a context-adaptive medical checklist through technology probe approach. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*, ACM, New York, NY, USA, 57–68. DOI:https://doi.org/10.1145/3064663.3064715
- [25] Pei-Yi Kuo, Rajiv Saran, Marissa Argentina, Michael Heung, Jennifer L. Bragg-Gresham, Dinesh Chatoth, Brenda Gillespie, Sarah Krein, Rebecca Wingard, Kai Zheng, and Tiffany C. Veinot. 2019. Development of a checklist for the prevention of intradialytic hypotension in hemodialysis care: Design considerations based on activity theory. ACM, 642. DOI:https://doi.org/10.1145/3290605.3300872
- [26] Angela Lashoher, Eric B. Schneider, Catherine Juillard, Kent Stevens, Elizabeth Colantuoni, William R. Berry, Christina Bloem, Witaya Chadbunchachai, Satish Dharap, Sydney M. Dy, Gerald Dziekan, Russell L. Gruen, Jaymie A. Henry, Christina Huwer, Manjul Joshipura, Edward Kelley, Etienne Krug, Vineet Kumar, Patrick Kyamanywa, Alain Chichom Mefire, Marcos Musafir, Avery B. Nathens, Edouard Ngendahayo, Thai Son Nguyen, Nobhojit Roy, Peter J. Pronovost, Irum Qumar Khan, Junaid Abdul Razzak, Andrés M. Rubiano, James A. Turner, Mathew Varghese, Rimma Zakirova, and Charles Mock. 2017. Implementation of the World Health Organization Trauma Care Checklist Program in 11 centers across multiple economic strata: Effect on care process measures. *World J Surg* 41, 4 (April 2017), 954–962. DOI:https://doi.org/10.1007/s00268-016-3759-8
- [27] Shauna M. Levy, Casey E. Senter, Russell B. Hawkins, Jane Y. Zhao, Kaitlin Doody, Lillian S. Kao, Kevin P. Lally, and Kuojen Tsao. 2012. Implementing a surgical checklist: more than checking a box. *Surgery* 152, 3 (September 2012), 331–336. DOI:https://doi.org/10.1016/j.surg.2012.05.034
- [28] Lorelei Lingard, Glenn Regehr, Beverley Orser, Richard Reznick, G. Ross Baker, Diane Doran, Sherry Espin, John Bohnen, and Sarah Whyte. 2008. Evaluation of a preoperative checklist and team briefing among surgeons, nurses, and anesthesiologists to reduce failures in communication. *Arch Surg* 143, 1 (January 2008), 12–17; discussion 18. DOI:https://doi.org/10.1001/archsurg.2007.21
- [29] Mark K. Lyons. 2010. Eight-year experience with a neurosurgical checklist. *Am J Med Qual* 25, 4 (August 2010), 285–288. DOI:https://doi.org/10.1177/1062860610363305
- [30] Oscar E. C. van Maarseveen, Wietske H. W. Ham, Nils L. M. van de Ven, Tim F. F. Saris, and Luke P. H. Leenen. 2019. Effects of the application of a checklist during trauma resuscitations on ATLS adherence, team performance, and patient-related outcomes: a systematic review. *Eur J Trauma Emerg Surg* (August 2019). DOI:https://doi.org/10.1007/s00068-019-01181-7
- [31] Tahrin Mahmood, Faizal Haji, Rita Damignani, Darius Bagli, Adam Dubrowski, Julian Manzone, Judy Truong, Robert Martin, and Maria Mylopoulos. 2015. Compliance does not mean quality: an in-depth analysis of the safe surgery checklist at a tertiary care health facility. *Am J Med Qual* 30, 2 (April 2015), 191. DOI:https://doi.org/10.1177/1062860614567830
- [32] Rajshri Mainthia, Timothy Lockney, Alexandr Zotov, Daniel J. France, Marc Bennett, Paul J. St Jacques, William Furman, Stephanie Randa, Nancye Feistritzer, Roland Eavey, Susie Leming-Lee, and Shilo Anders. 2012. Novel use of electronic whiteboard in the operating room increases surgical team compliance with pre-incision safety practices. *Surgery* 151, 5 (May 2012), 660–666. DOI:https://doi.org/10.1016/j.surg.2011.12.005
- [33] Carsten S. Østerlund. 2008. Documents in place: Demarcating places for collaboration in healthcare settings. *Comput. Supported Coop. Work* 17, 2–3 (April 2008), 195–225. DOI:https://doi.org/10.1007/s10606-007-9064-1
- [34] Sun Young Park and Yunan Chen. 2012. Adaptation as design: learning from an EMR deployment study. ACM, 2097–2106. DOI:https://doi.org/10.1145/2207676.2208361

- [35] Samantha E. Parsons, Elizabeth A. Carter, Lauren J. Waterhouse, Jennifer Fritzeen, Deirdre C. Kelleher, Karen J. O'connell, Aleksandra Sarcevic, Kelley M. Baker, Erik Nelson, Nicole E. Werner, Deborah A. Boehm-Davis, and Randall S. Burd. 2014. Improving ATLS performance in simulated pediatric trauma resuscitation using a checklist. *Ann. Surg.* 259, 4 (April 2014), 807–813.
DOI:<https://doi.org/10.1097/SLA.0000000000000259>
- [36] Janki Patel, Kamran Ahmed, Khurshid A. Guru, Fahd Khan, Howard Marsh, Mohammed Shamim Khan, and Prokar Dasgupta. 2014. An overview of the use and implementation of checklists in surgical specialities – A systematic review. *International Journal of Surgery* 12, 12 (December 2014), 1317–1323. DOI:<https://doi.org/10.1016/j.ijssu.2014.10.031>
- [37] Kathleen H. Pine and Melissa Mazmanian. 2014. Institutional logics of the EMR and the problem of “Perfect” but inaccurate accounts. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*, ACM, New York, NY, USA, 283–294.
DOI:<https://doi.org/10.1145/2531602.2531652>
- [38] George Ramsay, Alex B. Haynes, Stuart R. Lipsitz, Ivan M. Solsky, Jason Leitch, Atul A. Gawande, and Manoj Kumar. 2019. Reducing surgical mortality in Scotland by use of the WHO Surgical Safety Checklist. *BJS (British Journal of Surgery)* 106, 8 (2019), 1005–1011.
DOI:<https://doi.org/10.1002/bjs.11151>
- [39] Bradley N. Reames, Robert W. Krell, Darrell A. Campbell, and Justin B. Dimick. 2015. A checklist-based intervention to improve surgical outcomes in Michigan: evaluation of the Keystone Surgery program. *JAMA Surg* 150, 3 (March 2015), 208–215.
DOI:<https://doi.org/10.1001/jamasurg.2014.2873>
- [40] Madhu C. Reddy, Paul Dourish, and Wanda Pratt. 2001. Coordinating heterogeneous work: Information and representation in medical care. In *ECSCW 2001: Proceedings of the Seventh European Conference on Computer Supported Cooperative Work 16–20 September 2001, Bonn, Germany*. Springer Netherlands, Dordrecht, 239–258.
DOI:https://doi.org/10.1007/0-306-48019-0_13
- [41] Madhu C. Reddy, Paul Dourish, and Wanda Pratt. 2006. Temporality in medical work: Time also matters. *Comput Supported Coop Work* 15, 1 (February 2006), 29–53.
DOI:<https://doi.org/10.1007/s10606-005-9010-z>
- [42] Willem R. Spanjersberg, Engelbert A. Bergs, Nino Mushkudiani, Markus Klimek, and Inger B. Schipper. 2009. Protocol compliance and time management in blunt trauma resuscitation. *Emerg Med J* 26, 1 (January 2009), 23–27.
DOI:<https://doi.org/10.1136/emj.2008.058073>
- [43] Eric A. Sparks, Hania Wehbe-Janek, Rebecca L. Johnson, W. Roy Smythe, and Harry T. Papaconstantinou. 2013. Surgical Safety Checklist compliance: a job done poorly! *J. Am. Coll. Surg.* 217, 5 (November 2013), 867–873.e1–3.
DOI:<https://doi.org/10.1016/j.jamcollsurg.2013.07.393>
- [44] Carla Swarczinski and Phyllis Graham. 1990. From ICU to rehabilitation: a checklist to ease the transition for the spinal cord injured. *J Neurosci Nurs* 22, 2 (April 1990), 89–91.
- [45] David R. Urbach, Anand Govindarajan, Refik Saskin, Andrew S. Wilton, and Nancy N. Baxter. 2014. Introduction of surgical safety checklists in Ontario, Canada. *New England Journal of Medicine* 370, 11 (March 2014), 1029–1038.
DOI:<https://doi.org/10.1056/NEJMsa1308261>
- [46] Eefje N. de Vries, Hubert A. Prins, Rogier M. P. H. Crolla, Adriaan J. den Outer, George van Anel, Sven H. van Helden, Wolfgang S. Schlack, M. Agnès van Putten, Dirk J. Gouma, Marcel G. W. Dijkgraaf, Susanne M. Smorenburg, Marja A. Boermeester, and SURPASS Collaborative Group. 2010. Effect of a comprehensive surgical safety system on patient outcomes. *N. Engl. J. Med.* 363, 20 (November 2010), 1928–1937.
DOI:<https://doi.org/10.1056/NEJMsa0911535>
- [47] Joel S. Warm, Raja Parasuraman, and Gerald Matthews. 2008. Vigilance requires hard mental work and is stressful. *Hum Factors* 50, 3 (June 2008), 433–441.
DOI:<https://doi.org/10.1518/001872008X312152>
- [48] Alan M. Wolff, Sally A. Taylor, and Janette F. McCabe. 2004. Using checklists and reminders in clinical pathways to improve hospital inpatient care. *Med. J. Aust.* 181, 8 (October 2004), 428–431.
- [49] John E. Ziewacz, Sigurd H. Berven, Valli P. Mummaneni, Tsung-Hsi Tu, Olaolu C. Akinbo, Russ Lyon, and Praveen V. Mummaneni. 2012. The design, development, and implementation of a checklist for intraoperative neuromonitoring changes. *Neurosurgical Focus* 33, 5 (November 2012), E11.
DOI:<https://doi.org/10.3171/2012.9.FOCUS12263>