Toward Improving Surgical Outcomes by Incorporating Cognitive Load Measurement into Process-Driven Guidance

George S. Avrunin
University of Massachusetts
Amherst, Massachusetts, USA
avrunin@cs.umass.edu

Lori A. Clarke
University of Massachusetts
Amherst, Massachusetts, USA
clarke@cs.umass.edu

Heather M. Conboy
University of Massachusetts
Amherst, Massachusetts, USA
hconboy@cs.umass.edu

Leon J. Osterweil
University of Massachusetts
Amherst, Massachusetts, USA
ljo@cs.umass.edu

Roger D. Dias
Brigham and Women's Hospital
Boston, Massachusetts, USA
rdias@bwh.harvard.edu

Steven J. Yule
Harvard Medical School
Boston, Massachusetts, USA
SYULE@bwh.harvard.edu

Julian M. Goldman
Massachusetts General Hospital
Cambridge, Massachusetts, USA
JMGOLDMAN@mgh.harvard.edu

Marco A. Zenati
Harvard Medical School
Boston, Massachusetts, USA
Marco_Zenati@hms.harvard.edu

ABSTRACT

This paper summarizes the accomplishments and recent directions of our medical safety project. Our process-based approach uses a detailed, rigorously-defined, and carefully validated process model to provide a dynamically updated, context-aware and thus, "Smart" Checklist to help process performers understand and manage their pending tasks [7]. This paper focuses on support for teams of performers, working independently as well as in close collaboration, in stressful situations that are life critical. Our recent work has three main thrusts: provide effective real-time guidance for closely collaborating teams; develop and evaluate techniques for measuring cognitive load based on biometric observations and human surveys; and, using these measurements plus analysis and discrete event process simulation, predict cognitive load throughout the process model and propose process modifications to help performers better manage high cognitive load situations.

This project is a collaboration among software engineers, surgical team members, human factors researchers, and medical equipment instrumentation experts. Experimental prototype capabilities are being built and evaluated based upon process models of two cardiovascular surgery processes, Aortic Valve Replacement (AVR) and Coronary Artery Bypass Grafting (CABG). In this paper we describe our approach for each of the three research thrusts by illustrating our work for heparinization, a common subprocess of both AVR and CABG. Heparinization is a high-risk error-prone procedure that involves complex team interactions and thus highlights the importance of this work for improving patient outcomes.

CCS CONCEPTS

• **Human-centered computing** → *Computer supported cooperative work*; • **Software and its engineering** → *Integrated and visual development environments*;

KEYWORDS

Process modeling, cognitive load, checklists, simulation, surgical data science, surgical patient safety, augmented cognition

1 INTRODUCTION

This paper describes progress to date on our process-based medical safety project. This project is aimed at using precise, rigorously-defined, and detailed process models to improve medical outcomes by providing timely, context-aware guidance to coordinating teams, especially during periods of high cognitive load. To evaluate our approach, we are working with medical personnel to study and model cardiovascular surgery.

Cardiovascular surgery is a life-critical procedure that demands precise coordination of highly skilled teams of humans as well as the timely application of complex equipment. Unfortunately, it is also an error prone process, averaging about four errors an hour during the four or more hours of surgery [35]. Thus, it is an important example for which to develop, apply and evaluate our software-directed, process-based guidance approach to reducing errors.

Building upon our previous work, our current focus is threefold:

- How to provide guidance and support to teams of performers, so that each team has a clear picture of their process steps

and current context as well as understands the process state of the other teams they must interact with;

- How to measure and evaluate cognitive load so that we can predict more accurately the cognitive demand of the various process performers during a wide range of contexts;
- How to incorporate cognitive load information into the process model to help process performers handle high load situations better; for example, performers are encouraged to limit interruptions and noise during times of high cognitive load, and know when intra- and inter-team communication can best be received.

Cardiovascular surgery requires high levels of both technical skills (activities like tying a knot, applying a clamp, etc.) and non-technical ones involving situational awareness, communication, and teamwork. Both are critical to achieving good outcomes. In this paper, our focus is primarily on providing support for these non-technical aspects of the process (e.g., [13]).

Central to our approach is a cardiovascular surgical team process model: a precisely-defined software model that provides a detailed hierarchical view of how all the members of the overall surgical team do their work, and how the four specialty subteams (Surgery, Perfusion, Anesthesiology, and Nursing) coordinate with each other as well as with a suite of medical devices such as a cardiopulmonary bypass pump (also known as a heart-lung machine), a lung ventilator, and patient monitors. This model captures the normative (or usual) situations as well the non-normative (or exceptional) situations that can be expected to arise, as well as details of how these situations are to be handled. The model addresses, among other things, how each member of each specialty subteam performs their individual activities (or steps) and how team members interact with each other as well as with pertinent medical devices and software applications. The hierarchical nature of the model supports providing a high level view of team coordination as well as decompositions that support arbitrarily detailed specification of how process performers carry out their steps.

These models are being elicited, and iteratively improved, by interviewing surgical domain experts, by observing process performers, by consulting medical literature, and by codifying best practices. As described in earlier work, the models have been iteratively refined and improved by applying automated analyses (e.g., [4]) to identify process defects and vulnerabilities that domain experts believe to be hazardous, and to remove these defects and reduce these vulnerabilities so that the model can then be better trusted to guide the actual performance of surgery.

We are evaluating our approach by focusing on two surgical team processes, Aortic Valve Replacement (AVR) and Coronary Artery Bypass Grafting (CABG). For each of these processes, we have elicited precise and detailed models of several of their key high-level steps, as performed at the Veterans Affairs Boston Healthcare System (VABHS), and examples of process defects and vulnerabilities that are of particular concern to surgical domain experts. The surgical team process models are written in the Little-JIL process modeling language [6]. This language supports precise specification of complex semantics such as concurrency, exception management, human-directed choices, and resource management. In addition, the language has rigorously-defined semantics that can be used to

formally analyze these models before they are deployed to monitor progress and provide situational awareness during actual surgeries.

In the next section of this paper, we describe our current work on these three aspects of our approach. We then present related work and end with a discussion about the status of our project and future directions.

2 APPROACH

The members of an overall surgical team perform cardiovascular surgery in a high-demand and high-risk environment requiring simultaneous processing of large amounts of information (e.g., [8, 11, 37]). High demands imposed by surgical tasks may at times strain the cognitive capacity of process performers, perhaps leading to cognitive overload [33], which may impact performance negatively, increasing the risk of patient harm. Successful outcomes require the ability to deal with challenging situations, such as exceptional situations that may arise even as other exceptional situations are being dealt with. Members of the subteams must keep track of what their own team needs to do, but must also maintain "situational awareness" of the process steps of the other subteams. Patient safety requires that errors be avoided, and their effects mitigated when they occur.

The Multi-Team Smart Checklist draws upon the process model to provide real-time guidance and situational awareness by presenting the state of this complex concurrent process sufficiently clearly that each process performer can see how each subteam is doing its work, including what it has done so far and what the (possible) next steps will be. This work builds upon previous work (e.g., [7]) that described a Single-Team view that provides a sequential view of the process execution state. In this current work the Multi-Team view extends this Single-Team view by displaying the steps of each subteam, as well as indicating the ways in which one subteam's steps interact with and depend on the steps of the others. The Multi-Team Smart Checklist also uses the process model to provide a view of upcoming situations, helping to prepare process performers for impending periods of high cognitive load.

Another key focus of our work will be to identify the situations that produce the high cognitive load that increases the likelihood of errors. To that end, our research team has been gathering biometric and survey data from actual cardiovascular surgery process performers during and immediately after surgery. We are observing that certain process steps are more likely to cause high load, but also that the contexts in which these steps are performed have a strong effect on cognitive load as well. Patient condition inferred from data gathered and synthesized from medical devices can create such contexts, as can the recent occurrence of worrisome events, as well as data taken from the skill levels of key surgical team members and the states of surgical devices such as cardiopulmonary bypass pumps. In addition, analyzing the structure of our process models, such as identifying when an exception, or a nest of exceptions, is being handled can be used to predict cognitive load, as can analysis of the path that the process has traversed.

Understanding of the effects that each of these contextual factors have on cognitive load will also be used to create suggestions for how to avoid these situations or mitigate them. Such understandings might be used, for example, to suggest choices when different

options are available, post broadcast information to assure better coordination, or suggest making changes in staffing (e.g. bringing in additional high-skill personnel). We will use speculative execution of these process models to determine when and how particularly challenging situations might occur in order to advise the surgical team of their imminence, and suggest how to avoid them or mitigate their impact on team performance.

Cardiovascular surgery provides a significant test case for our approach. Our process model of the CABG process, parts of which are still being elaborated, is written in Little-JIL and consists of 14 high-level steps whose decomposition consists of about 200 steps. This process model specifies the recommended process, including the normative scenarios as well as the non-normative scenarios. It was elicited through observations of, and multiple interviews with, members of each of the four subteams at a teaching hospital of Harvard Medical School, many of whom also provided us with best practice guidelines and training documents. We expect that these processes may be performed differently at different hospitals, and look forward to determining how well our current models could be adapted to represent practices at other locations. But we believe that the general approach described here for VABHS will also prove useful in other locations and, indeed, to a wide range of other critical processes involving the interaction of teams of humans, devices, and software applications.

2.1 Context-Aware, Dynamic Guidance Using Smart Checklists

Checklists that guide humans through the performance of processes in various domains have been shown to reduce errors (e.g., [14]). Typical checklists, however, can be inadequate in guiding humans through the complex aspects of intricate processes such as surgical team processes, which are known to be inherently error prone (e.g., [34]). We are developing and evaluating a framework that dynamically generates *Smart Checklist* user interfaces [7] that provide context-aware guidance to humans as they are performing a real-world process. This framework updates the Smart Checklists, which visualize the process execution state, by monitoring real-time process execution events and then matching them against sequences of events specified in the validated process models. In previous work, we presented an initial prototype of a Single-Team Smart Checklist which provides an individual view for a single subteam that displays their process execution state, including previously performed steps; step(s) currently being performed; and potential future steps for alternative ways forward. Here we present a design for a Multi-Team Smart Checklist that, in addition to single-team views, provides a shared view for multiple subteams that displays the complicated interactions among subteams.

In order to update the Smart Checklists, our framework must recognize events that occur in the actual process execution. For steps completed by human process performers, this information can be reported to the framework by the agent interacting with the Smart Checklist, by a scribe tracking the process in the operating room, or possibly even analyzing video or audio streams. The framework, however, must also gather information from a number of complex medical devices. In our project, this is handled

by the OpenICE (Open Integrated Clinical Environment) [2] open-source implementation of the ASTM F2761-09(13) standard [3]. In our work OpenICE assumes responsibility for gaining access to device-generated data and makes it available to process subscribers. Some devices, like patient monitors and ventilators, monitor several hundred variables, including device settings, alarms, technical alerts, and internal device component status, in addition to values monitored from patients. Thus OpenICE also creates more abstract, higher-level information that combines and summarizes lower level data elements, where this data may even come from several different devices. This facilitates the incorporation into our process models of conditions or guard statements that may be based on complicated calculations where the mathematics may not be appropriate to include in the process model. Thus, complex calculations, perhaps based on data from diverse devices are done by OpenICE, and are presented as a simple variable to the process model. This approach has been used, for instance, to replace the monitoring and integration of lung gas flow values with a single "patient-breath-finished" Boolean variable [1]. Additionally, OpenICE could be used to integrate with medical software applications such as electronic health records.

Multi-Team Smart Checklists need to guide surgical teams through processes that often involve complex and overlapping interactions. Since multiple subteams are often concurrently performing their steps, those subteams often need to communicate shared data and correctly synchronize or order the steps. For a given subteam, some of their steps may need to be performed individually and others may need to be performed collaboratively with at least one other subteam. Additionally, a subteam may need to go on standby while other subteams complete some steps before continuing on to perform their own steps. We therefore designed the Multi-Team view so that it can display information about the multiple subteams and their key shared data. We also designed this view to allow the overall surgical team to selectively show a Single-Team view for each subteam, showing a listing of that subteam's steps as well as any other subteam's steps that they are waiting on. The Multi-Team view must keep those Single-Team views synchronized as the different subteams progress through the process.

We illustrate this here with a part of the **Heparinization** subprocess that is common to both CABG and AVR surgery. Heparinization is a complex task that is critically important. It incorporates the detection and response to a number of different contingencies that can add significantly to the cognitive load of its performers. It is also of critical importance because incorrect performance can create the serious risk of stroke during the cardiovascular surgery, due to clots that form when the patient's blood comes in contact with the highly thrombogenic plastic cannulae (specialized tubing) that have been inserted directly into the bloodstream, or as blood circulates through the cardiopulmonary bypass pump. To reduce this risk, the anticoagulant drug heparin is administered to reduce the tendency of the blood to clot, as measured by the *activated clotting time* (ACT). So a key phase of standard CABG or AVR surgery is the administration of heparin before inserting the cannulae and initiating cardiopulmonary bypass. This is a complex process, involving primarily the Surgery, Anesthesiology, and Perfusion specialty subteams. If the initial dose of heparin does

not provide sufficient anticoagulation, a series of additional measures based on best practices [12, 32] will be applied, with the ACT checked after each is tried. If none of these is successful, the surgeon may decide to switch to an "off-pump" procedure (in the case of CABG) or to abort the procedure altogether.

Figure 1 shows a Multi-Team Smart Checklist to guide the overall surgical team, consisting of the Anesthesiology (denoted by 'A'), Nursing (denoted by 'N'), Perfusion (denoted by 'P'), and Surgery (denoted by 'S') subteams, as they execute the **Perform isolated CABG** process. ("Isolated" means that it is not performed in combination with another procedure such as AVR.) Specifically, the Anesthesiology, Perfusion, and Surgery subteams are working together on the **Perform heparinization** subprocess (the subteams are denoted by the prefix 'A,P,S:' and the green background indicates this subprocess is currently being performed by those subteams). This Smart Checklist supports context-awareness in several ways, including the team information (shown by the team tabs along the top and the team table on the top left), the patient information (shown in the top middle), and the process execution state (shown on the top right and the bottom). The Smart Checklist can be customized by selecting which information (or state) should be shown and how to display that information. For instance, any step name that starts with perform (e.g., **Perform isolated CABG**) can be abbreviated (e.g., **Isolated CABG**).

In this figure, the overall surgical team decided to show a Multi-Team view (by selecting the tab labeled with that view) and then selecting the Anesthesiology, Perfusion, and Surgery subteams (by selecting the items in the team table for those subteams), since these teams need to coordinate their progress during the **Heparinization** subprocess. The patient information includes some selected patient identifiers, real-time vital sign data provided by OpenICE, and key subprocess-related data such as the *target ACT* that heparinization is aiming to achieve. The process execution state includes the high-level subprocess listing (shown on the top right). Additionally, this state includes the **Heparinization** subprocess listing (shown on the bottom), consisting of Single-Team views for the Anesthesiology subteam (shown on the bottom left), the Perfusion subteam (shown on the bottom middle), and the Surgery subteam (shown on the bottom right).

For the **Heparinization** subprocess, a recommended dose of heparin is calculated before the surgery according to the guidelines [32]. To begin this subprocess, that dose is physically administered to the patient by the anesthesiologist. After waiting at least **3** minutes, the anesthesiologist draws a blood sample and the perfusionist tests its ACT. In each of the three Single-Team views that are shown in this figure, the Perfusion team previously completed step **Test blood sample for ACT** (indicated by the gray backgrounds along with the green checkmark icons and the timestamps). The Perfusion subteam's view shows that this subteam is currently performing step **Confirm target ACT achieved** (indicated by the green background along with the green checkmark and red X buttons). For the current step, the subteam may consult with the Anesthesiology and Surgery subteams to decide whether or not the target ACT has been achieved (indicated by the prefix 'P,(A,S):'). The Anesthesiology and Surgery subteams' views also show that these two subteams may be consulted about this decision (indicated

by the green backgrounds). For the normative situation, the Perfusion subteam would decide that the target ACT has been achieved and this team should click on the green checkmark button. This situation has the potential future step **If decision to use pump sucker = YES, turn on pump sucker** (the white background indicates that this is a future step). For the exceptional situation where the problem *Low ACT* was identified, the Perfusion subteam should click on the red X button to report the identified problem. This situation has the potential future step **If low ACT, follow heparin resistance protocol**. Additionally, the Perfusion subteam could decide to document their clinical notes by clicking on the notepad button to bring up a dialog box and then type in the notes.

Figure 2 depicts the process execution state just after the Anesthesiology, Perfusion, and Surgery subteams have completed the step **Confirm target ACT achieved** and reported that the problem *Low ACT* was identified (shown towards the middle of the **Heparinization** process listing with a gray background, a red X icon, and timestamp "15:49"). To address that problem, the Anesthesiology, Perfusion, and Surgery subteams have started to perform **If low ACT, follow heparin resistance protocol** (shown with a green background). In the Perfusion subteam's view, the **Gather Anesthesiology, Perfusion, and Surgery team leaders** is shown with a green background and a green checkmark button, indicating that the Perfusion subteam needs to perform this step. On the other hand, the Anesthesiology subteam's view shows that Anesthesiologist is waiting for the Perfusionist to perform the step (the yellow background and grayed out text indicates the subteam is waiting for some other process performer to complete this step). For the exceptional situation where the problem *Low ACT* is being addressed, the potential future step is **Treat heparin resistance** that specifies that the subteams try several alternatives in the listed order, trying to successfully achieve the target ACT. If any of these alternatives succeeds, the potential future step is **If decision to use pump sucker = YES, turn on pump sucker**. If none of the alternatives succeed, the overall surgical team can either attempt an alternative procedure (an "off-bypass" approach that does not use the pump, which is not possible for AVR) or abort the surgery.

The proposed Smart Checklist design has been very favorably received by focus groups of cardiovascular surgical team members. The team members could see the potential of such checklists to improve team training and reduce procedural errors by their inclusion of details about key shared data, necessary team communication, and exceptional situations. The focus groups suggested that each subteam have its own monitor to display its own perspective on an ongoing process. They also suggested that all subteams have a shared monitor to display all interactions during key process steps such as heparinization. The focus groups also indicated that certain process steps, either by definition or because of their context, have a high cognitive load and thus are more likely to lead to procedural errors. The next two subsections describe how to measure or predict the cognitive load for such steps and how the Smart Checklists could modify the guidance provided based on those cognitive loads. Our next evaluation of these checklists will entail live human simulations aimed at guiding the entire team, or part of the team, through various normative and exceptional scenarios, drawn from recordings to be described in the next subsection.
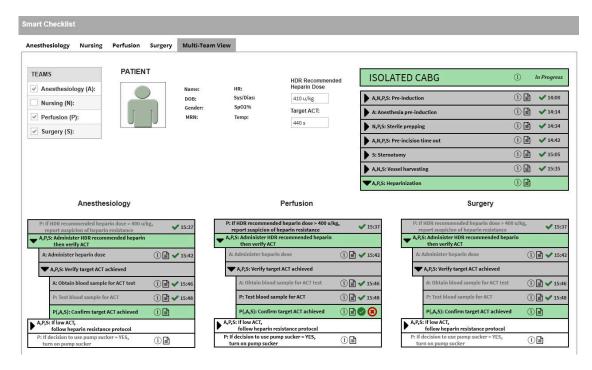
Figure 1: Multi-Team Smart Checklist to guide the Anesthesiology, Perfusion, and Surgery subteams through the Perform isolated CABG process
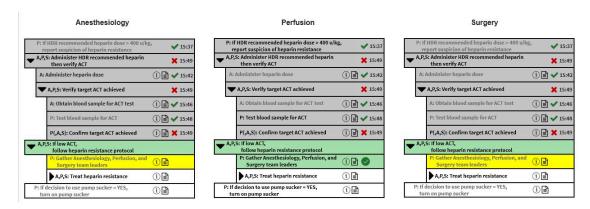

Figure 2: Updates to the Multi-Team Smart Checklist after the Perfusion team completes step Confirm target ACT achieved and reports that the problem *Low ACT* needs to be addressed

2.2 Dynamically Measuring Cognitive Load Using Heart Rate Variability Analysis and Surveys

It is well established that errors are more likely when process performers are under high cognitive load [13, 34]. The process guidance and situational awareness provided by the Multi-Team Smart Checklist are expected to help reduce the likelihood of certain kinds of errors directly, by reminding surgical team members of the steps that need to be performed in a particular context, and indirectly, by providing additional situational awareness. But the Smart Checklist

could also be used to directly reduce the cognitive load of team members by, for instance, indicating periods when a particular team member should not be interrupted with non-emergency communications or even calling for additional resources when they would help a team member with a particularly complex situation. To achieve these benefits, however, we need a way of measuring the cognitive load on individual team members in something close to real time.

A well-validated retrospective measure of (self-perceived) cognitive load in surgery is the SURG-TLX questionnaire [36], based on the NASA-TLX (Task Load Index) [15]. But this questionnaire

is completed after the surgery and does not provide information that can be used in guiding team members during the surgery. After a careful review [9] and validation against SURG-TLX [10], we have chosen heart rate variability (HRV) as an objective measure of cognitive load that can be obtained in near real time. We are currently combining video and audio recordings of CABG and AVR surgeries with measures of inter-beat (R-R) intervals captured by a smartphone app connected to heart rate sensors worn by the lead member of each subteam. The measure of cognitive load can be obtained from spectral analysis of this data. We are also using the recordings to associate cognitive load with specific process steps in our cardiovascular surgery process models.

Our preliminary data suggest that this approach is promising and can begin to allow us to use real-time measures of cognitive load to identify situations in which particular team members are under high (or low) cognitive load and use this information to inform process guidance. We also hope to be able to associate high cognitive load with particular subprocesses and contexts.

2.3 Reducing Errors by Predicting Cognitive Load

Some of this information about the process performer's cognitive load can be used directly by the Smart Checklist. For instance, it is well-established that interruptions during periods of high cognitive load are very disruptive and substantially increase the load and the likelihood of errors. The Smart Checklist could be used to warn the members of other subteams that a particular subteam or subteams are under high cognitive load and communication with them should be limited to essential information. The Checklist could even request other subteams to minimize all communication and other noise during particularly high-load periods. Or the Checklist might provide more detailed guidance during periods of high cognitive load, so that performers are reminded of details that they might ordinarily be expected not to require.

But our long-term goal is to provide guidance that will, when possible, help re-direct the process execution away from high cognitive load situations. For instance, it might be possible to recognize that a period of especially high cognitive load is about to arise and request additional resources, such as advice from a senior colleague or call for the delivery of special equipment. One approach we are exploring is to use a process simulator to traverse forward from the current process state through the process model, exploring multiple paths when control flow alternatives are encountered. If we can associate appropriate levels of cognitive load to process steps in particular contexts, the simulator could estimate the cognitive load along each of these possible executions for a given number of steps, or a certain amount of time, into the future. If a critical level of cognitive load is predicted along such a path, or perhaps if an appropriate function of the cognitive loads predicted for all impending paths exceeds some criterion, the Checklist could propose steps to reduce or avoid that load. In the future, we might be able to guide execution by defining and monitoring an optimal level of cognitive load, similar to what has been proposed for aviation. Such an approach requires good estimates of cognitive load for steps in context, which we hope to begin to get from our measurements of cognitive load in actual surgeries, and information about measures

that might help reduce the load. That latter information would be derived from a careful process design effort involving the medical domain experts.

We are currently modifying our discrete event simulator to track cognitive load along a simulated process path. This approach needs to support considerations of the inherent load, for each subteam, of a particular step as well as the effect of various process contexts, e.g., entering or leaving the exception handlers that were described as the various responses to low ACT during the **Heparinization** subprocess. Similarly, it will be important to predict an increase in cognitive load during periods of the process where several subteams must coordinate and communicate effectively, with the cognitive load of the communicating subteams perhaps increasing nonlinearly with the number of subteams that must communicate.

Both the self-reported measures of cognitive load and our initial data from HRV measurements indicate that another important factor in a process performer's cognitive load is the expertise level of various team members. The VABHS is a teaching hospital, and so training of new personnel is typically part of surgeries. Our data indicate, not surprisingly, that the cognitive load on a senior process performer is significantly increased when that person is teaching as part of the process. But even when no explicit teaching is being performed, we have observed, for instance, that the surgeon's cognitive load is higher when working with a less experienced anesthesiologist. As part of the ROMEO resource management component of our simulator [28], we have the ability to quantify certain characteristics of the individual human process performers, and to use those quantities in computing cognitive load adjustment. ROMEO supports the specification of each resource, human and non-human, that is a candidate for participation in the performance of a process. Each specification contains a set of quantified attributes. At present this set includes experience level, and skill level, where the level of the process performer's skill can be specified for each of the steps of the process. But we expect our research to indicate the need to include additional attributes in our model so that we will be better able to predict cognitive load along simulated execution paths.

3 RELATED WORK

Medical processes are known to be inherently complex and error prone (e.g., [11, 21]). Paper checklists have been used in multiple domains, initially in aviation (e.g., [5]) and now in medicine (e.g., [14]), and shown to be able to significantly reduce errors. One objection to such checklists, however, is that they describe the normative situations but not the exceptional situations where a problem is identified and then must be addressed. Another objection to these checklists is that they are static and hence unable to make adjustments for context.

Including exceptional situations can be useful for novices as well as experts because these situations may rarely occur and how the identified problem is addressed may be complex, involving team communication, collaborative decision making, and subsequent exceptional situations. In the medical domain, some of the commonly occurring exceptional situations that are high risk for the patient have been standardized as crisis checklists (e.g., [39]). In our **Heparinization** subprocess model, the step **Follow heparin**

resistance protocol is based on a customized crisis checklist developed at the VABHS [18].

Electronic checklists (e.g., [16]) have been introduced to be able to visualize the process execution state and dynamically update that state based on the actual process execution events produced by the medical team and automated components such as medical devices and software applications. Some of these electronic checklists provide simple visualizations of the paper-based checklists. Other electronic checklists (e.g., [24, 29]) support complex visualizations of the process execution state including information about the medical team, the patient, the process steps, etc. Some of this work, most notably the Tracebook project [26], is also based on a process model. While these process-model based projects share our goals, the process notations that they use seem to us to lack the powerful semantics required (e.g exception management, resource specification, and procedural abstraction) to represent complex processes such as cardiac surgery with the necessary detail, precision, and clarity.

Some related work (e.g., [19]) uses cognitive informatics techniques to evaluate how the cognitive abilities of a single clinician or team affect the performance of medical processes, impacting healthcare outcomes. Such cognitive evaluation provides feedback to clinicians, suggesting how to better perform the processes and improve outcomes. In our work, we described how Smart Checklists can provide guidance to clinicians performing medical processes and described our efforts to measure and predict cognitive load contexts for those processes. As future work, we plan to investigate how to reduce procedural errors by using different cognitive load contexts to tailor the guidance we provide to clinicians.

Other related work on measuring cognitive load either during or immediately after surgery is discussed in detail in [9].

Discrete event simulation, driven by articulate models of processes, is a research area that has been pursued by a number of researchers for at least 20 years [20, 27, 38]. More recently discrete event simulations of processes has become more ambitious, incorporating the use of orthogonal specifications of the behaviors and characteristics of agents performing the steps articulated in the simulation model [28]. In previous work we have reported on early efforts to use this kind of mixed agent/activity simulation for medical processes [31]. The work reported here extends that previous work by incorporating representations of the cognitive load on agents as a key focus of the simulations.

4 CONCLUSIONS

Although 18 years have elapsed since the seminal U.S. Institute of Medicine report on medical errors and patient safety, healthcare safety still lags significantly behind other high-reliability organizations (e.g., [21, 23, 34]). To fill this safety gap, new medical error and patient safety management approaches are necessary.

Complex procedural healthcare (e.g. surgery, interventional radiology, interventional cardiology, etc.) is a high-consequence team-based sociotechnical system with critical requirements for communication and coordination (e.g., [13, 23, 34]). Contemporary sociotechnical systems research has moved away from the individual as the unit of cognitive analysis, and a new focus on the activity system (a group of human actors, their tools and environment) has

been proposed; this framework has been referred to as "distributed cognition" (e.g., [17, 30]).

Our project has taken an important step towards operationalization of many of these ideas. We have materialized surgical team coordination in the form of a precise, validated, rigorously-defined process model and have embedded observations about the loci of high cognitive load situations in the model. We are evaluating whether this information can be used to help surgical team process performers anticipate these situations, and to guide the participants through and/or around them. But much work remains to be done.

Thus, for example, we need to continue to improve our methods for recording human-initiated events and communications, to gain more complete access to medical device data, to sharpen our ability to coordinate process model state with the state of the actual process. We rely heavily on scribes to do this now, but automated approaches such as voice recognition should be explored, as well as enhancements to the Smart Checklist user interface, aimed at facilitating wider categories of human input. We are also exploring how best to specify and monitor adherence to timing constraints, and use them to help performers without annoying them.

Our approach is based upon iterative incremental improvement. Thus our models are increasingly broader, deeper and more accurate, and the properties, constraints, and hazards against which they are evaluated continue to grow in number and precision. Our observations are being sharpened and enhanced as we continue to gather data from monitoring increasing numbers of surgeries. We are similarly refining our measurements of cognitive load and the associations of increased load with particular steps and contexts (especially non-normative contexts) in the process model, and we are using discrete event simulation to sharpen and validate these associations.

Our work fits well with work on detailed process models that focus on the fine-grained activities of a single process performer, such as the surgeon [22]. For example, the contexts provided by our higher-level team activities should be useful in helping to sharpen and adapt models of these fine-grained activities, and we look forward to incorporating these details into our models as well.

Our work should also complement work in the new area of Surgical Data Science [25]. We expect that the progress being made in that area will yield events, insights, and measurements that will be useful in further iterative improvements to our models. Conversely we believe that our work should be of substantial benefit to the Surgical Data Science community by suggesting hierarchical frameworks within which to structure observed and recorded data streams. As noted above, context plays a very important role in understanding and supporting surgical processes. While it may be possible to create such contexts by inferring hierarchy and other forms of structure from raw surgical event streams, we believe that, in making such structure explicit, our work will provide a framework for creating useful contextual information from the surgical data event streams that are collected.

REFERENCES

[1] David Arney, Julian M. Goldman, Susan F. Whitehead, and Insup Lee. 2010. *Improving Patient Safety with X-Ray and Anesthesia Machine Ventilator Synchronization: A Medical Device Interoperability Case Study*. Springer, Berlin, Heidelberg, 96–109.

[2] Dave Arney, Jeff Plourde, and Julian Goldman. 2018. OpenICE medical device interoperability platform overview and requirement analysis. *Biomed. Eng./Biomedizinische Technik* 63, 1 (Feb. 2018), 39–48.

[3] ASTM International. 2009. *ASTM F2761-2009. Medical Devices and Medical Systems—Essential Safety Requirements for Equipment Comprising the Patient-Centric Integrated Clinical Environment (ICE), Part 1: General Requirements and Conceptual Model*.

[4] George S. Avrunin, Lori A. Clarke, Leon J. Osterweil, Stefan C. Christov, Bin Chen, Elizabeth A. Henneman, Philip L. Henneman, Lucinda Cassells, and Wilson Mertens. 2010. Experience modeling and analyzing medical processes: UMass/Baystate medical safety project overview. In *Proc. of the 1st ACM Int. Health Inform. Symp.* 316–325.

[5] Daniel Boorman. 2001. Today's electronic checklists reduce likelihood of crew errors and help prevent mishaps. *International Civil Aviation Organization Journal* 56, 1 (2001), 17–36.

[6] Aaron G. Cass, Barbara Staudt Lerner, Stanley M. Sutton, Jr., Eric K. McCall, Alexander Wise, and Leon J. Osterweil. 2000. Little-JIL/Juliette: A process definition language and interpreter. In *Proc. of the 22nd Int. Conf. on Softw. Eng.* ACM, New York, NY, USA, 754–757.

[7] Stefan C. Christov, Heather M. Conboy, Nancy Famigletti, George S. Avrunin, Lori A. Clarke, and Leon J. Osterweil. 2016. Smart Checklists to Improve Healthcare Outcomes. In *Proc. of the Int. Workshop on Softw. Eng. in Healthcare Syst. (SEHS '16)*. ACM, New York, NY, USA, 54–57.

[8] Heather M. Conboy, George S. Avrunin, Lori A. Clarke, Leon J. Osterweil, Julian M. Goldman, Steven J. Yule, Marco A. Zenati, and Stefan C. Christov. 2017. Cognitive support during high-consequence episodes of care in cardiovascular surgery. In *2017 IEEE Conf. on Cognitive and Computational Aspects of Situation Management (CogSIMA)*. 1–3.

[9] Roger D. Dias, M. C. Ngo-Howard, M. T. Boskowski, Marco A. Zenati, and Steven J. Yule. 2018. Systematic Review of Measurement Tools to Assess Surgeons' Intraoperative Cognitive Workload. *Br J Surg* (2018). Published online.

[10] Roger D. Dias, Marco A. Zenati, Gabany Jennifer M., Heather M. Conboy, Lori A. Clarke, George S. Avrunin, Leon J. Osterweil, Julian M. Goldman, and Steven J. Yule. 2018. Embedding Real-Time Measure of Surgeons' Cognitive Load into Cardiac Surgery Process Modeling. *Academic Surgical Congress* (2018).

[11] Jonathan F. Finks, Nicholas H. Osborne, and John D. Birkmeyer. 2011. Trends in Hospital Volume and Operative Mortality for High-Risk Surgery. *N Engl J of Med* 364, 22 (2011), 2128–2137.

[12] Alan Finley and Charles Greenberg. 2013. Heparin Sensitivity and Resistance: Management During Cardiopulmonary Bypass. *Anesth Analg* 116, 6 (2013), 1210–1222.

[13] Rhona H. Flin, George G. Youngson, and Steven Yule (Eds.). 2015. *Enhancing surgical performance: a primer in non-technical skills*. Boca Raton.

[14] Brigette M. Hales and Peter J. Pronovost. 2006. The checklist: a tool for error management and performance improvement. *J Crit Care* 21 (2006), 231–235.

[15] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology* 52 (1988), 139–183.

[16] L. A. Hassell, A. V. Parwani, L. Weiss, M. A. Jones, and J. Y. 2010. Challenges and opportunities in the adoption of College of American Pathologists checklists in electronic format: perspectives and experience of reporting pathology protocols project (RPP2) participant laboratories. *Archives of Pathology and Laboratory Medicine* 134, 8 (2010), 1152–1159.

[17] Brian Hazlehurst, Paul N. Gorman, and Carmit K. McMullen. 2008. Distributed cognition: An alternative model of cognition for medical informatics. *Int J of Med Inform* 77, 4 (2008), 226–234.

[18] S. A. Hirji, C. Tarola, H. Amirfarzan, S. J. Yule, J. M. Gabany, R. D. Dias, J. A. Quin, M. Haime, and M. A. Zenati. 2017. Utility and Feasibility of Intra- and Postoperative Crisis Management Checklists in Cardiac Surgery. In *Proceedings, 14th Annual Society of Thoracic Surgeons Critical Care Conference*.

[19] Thomas G. Kannampallil, Joanna Abraham, and Vimla L. Patel. 2016. Methodological framework for evaluating clinical processes: A cognitive informatics perspective. *Journal of Biomedical Informatics* 64 (2016), 342 – 351.

[20] Marc I. Kellner, Raymond J. Madachy, and David Raffo. 1999. Software process simulation modeling: Why? What? How? *Journal of Systems and Software* 46, 2-3 (1999), 91–105.

[21] L. T. Kohn, J. M. Corrigan, and M. S. Donaldson (Eds.). 1999. *To Err Is Human: Building a Safer Health System*. National Academies Press.

[22] Florent Lalys and Pierre Jannin. 2014. Surgical Process Modelling: A Review. *Int J of Computer-Assisted Radiology and Surg* 9 (2014), 495–511.

[23] Nancy Leveson, Aubrey Samost, Sidney Dekker, Stan Finkelstein, and Jai Raman. 2017. A Systems Approach to Analyzing and Preventing Hospital Adverse Events. *To appear in J Patient Saf* (2017).

[24] Fitzgerald M, Cameron P, Mackenzie C, and et al. 2011. Trauma resuscitation errors and computer-assisted decision support. *Archives of Surgery* 146, 2 (2011), 218–225.

[25] Lena Maier-Hein, Swaroop S. Vedula, Stefanie Speidel, Nassir Navab, Ron Kikinis, Adrian Park, Matthias Eisenmann, Hubertus Feussner, Germain Forestier, Stamatia Giannarou, Makoto Hashizume, Darko Katic, Hannes Kenngott, Michael Kranzfelder, Anand Malpani, Keno März, Thomas Neumuth, Nicolas Padoy, Carla Pugh, Nicolai Schoch, Danail Stoyanov, Russell Taylor, Martin Wagner, Gregory D. Hager, and Pierre Jannin. 2017. Surgical data science for next-generation interventions. *Nature Biomedical Engineering* 1 (Sept. 2017), 691–696.

[26] Shan Nan, Pieter Van Gorp, Xudong Lu, Uzay Kaymak, Hendrikus Korsten, Richard Vdovjak, and Huilong Duan. 2017. A meta-model for computer executable dynamic clinical safety checklists. *BMC Medical Informatics and Decision Making* 17, 1 (12 Dec 2017), 170.

[27] David Raffo. 1998. Software Process Simulation Modeling (ProSim'98): Workshop Report. *Empirical Software Engineering* 3, 4 (1998), 407–412.

[28] M. S. Raunak and L. J. Osterweil. 2013. Resource Management for Complex, Dynamic Environments. *IEEE Transactions on Software Engineering* 39, 3 (March 2013), 384–402.

[29] Jeffrey Robbins. 2011. Hospital Checklists. *Critical Care Nursing Quarterly* 34, 2 (2011).

[30] N. Saberi, M. Mahvash, and M. A. Zenati. 2015. An artificial system for selecting the optimal surgical team. In *2015 37th Annual Int. Conf. of the IEEE Eng. in Medicine and Biology Society (EMBC)*. 218–221.

[31] Seung Yeob Shin, Yuriy Brun, and Leon J. Osterweil. 2016. Specification and Analysis of Human-intensive System Resource-utilization Policies. In *Proceedings of the International Workshop on Software Engineering in Healthcare Systems (SEHS '16)*. ACM, New York, NY, USA, 8–14.

[32] Linda Shore-Lesserson, Robert A. Baker, Victor A. Ferraris, Philip E. Greilich, David Fitzgerald, Philip Roman, and John W. Hammon. 2018. The Society of Thoracic Surgeons, The Society of Cardiovascular Anesthesiologists, and the American Society of ExtraCorporeal Technology: Clinical Practice Guidelines—Anticoagulation During CardioPulmonary Bypass. *The Annals of Thoracis Surgery* 105 (2018), 650–62. Co-published in *Anestesia & Analgesia* and the *Journal of ExtraCorporeal Technology*.

[33] C. L. Tarola, J. A. Quin, M. E. Haime, J. M. Gabany, K. B. Taylor, K. B. Leissner, and M. A. Zenati. 2016. Computer-Assisted Process Modeling to Enhance Intraoperative Safety in Cardiac Surgery. *JAMA Surg* 151, 12 (2016), 1183–1186.

[34] J. A. Wahr, R. L. Prager, J. H. Abernathy III, E. A. Martinez, et al. 2013. Patient Safety in the Cardiac Operating Room: Human Factors and Teamwork: A Scientific Statement from the American Heart Association. *Circulation* 128, 10 (2013), 1139–69.

[35] D. A. Wiegmann, A. W. ElBardissi, and J. A. Dearani. 2007. Disruptions in surgical flow and their relationship to surgical errors: an exploratory investigation. *Surgery* 142, 5 (Nov. 2007), 658–65.

[36] M. R. Wilson, J. M. Poolton, N. Malhotra, K. Ngo, E. Bright, and R. S. W. Masters. 2011. Development and Validation of a Surgical Workload Measure: The Surgery Task Load Index (SURG-TLX). *World J Surg* 35, 9 (2011), 1961–1969.

[37] M. A. Zenati, M. Grumbine, R. Pieraccini, J. A. Gabany, J. Quin, M. Haime, and J. Goldman. 2015. A Novel Point-of-Care Intraoperative Vigilance System to Enhance Surgical Quality and Safety. In *NIH-IEEE 2015 Strategic Conf. on Healthcare Innovations and Point-of-Care Technologies*.

[38] He Zhang, Barbara A. Kitchenham, and Dietmar Pfahl. 2010. Software Process Simulation Modeling: An Extended Systematic Review. In *ICSP 2010*. 309–320.

[39] John E. Ziewacz, Alexander F. Arriaga, Angela M. Bader, William R. Berry, Lizabeth Edmondson, Judith M. Wong, Stuart R. Lipsitz, David L. Hepner, Sarah Peyre, Steven Nelson, Daniel J. Boorman, Douglas S. Smink, Stanley W. Ashley, and Atul A. Gawande. 2011. Crisis Checklists for the Operating Room: Development and Pilot Testing. *Journal of the American College of Surgeons* 213, 2 (2011), 212 – 217.e10.