SYSTEMATIC OBSERVATION IN HEALTH CARE: UTILITY AND LIMITATIONS OF A THREAT AND ERROR MANAGEMENT-BASED SAFETY AUDIT

Andrew Grose, M.D.| LUND UNIVERSITY

SYSTEMATIC OBSERVATION IN HEALTHCARE: UTILITIES AND LIMITATIONS OF A THREAT AND ERROR MANAGEMENT-BASED SAFETY AUDIT

Thesis/Project work submitted in partial fulfillment of the requirements for the MSc in Human Factors and System Safety

Andrew Grose, M.D.

Under supervision of Johan Bergström, PhD.

Lund 2018

ABSTRACT

Improving teamwork has become a major safety goal for healthcare organizations. Audit tools currently available are useful, but they remain inadequate because they are reactive and fail to provide context for "the interaction between people and the operational context (i.e., organizational, regulatory and environmental factors) within which people discharge their operational duties" (Maurino, 2005). Accurate and relevant information about real-world team behavior is theorized to confer the ability to address, through process design &/or training, significant issues which can then be re-assessed through repeat observations. In the mid-1990s, the Federal Aviation Administration (FAA) funded collaboration between the University of Texas and Continental Airlines to directly observe in-flight behaviors associated with safety and risk. This methodology, now known as the *Line Operation Safety Audit* (LOSA), was instrumental in developing the Threat and Error Management (TEM) model of cockpit work performance. In 2006, the FAA made TEM-based LOSA a "voluntary safety recommendation," and all major US commercial air carriers engage in this on a regular basis as a component of their safety management systems (FAA, 2006).

This thesis describes the adaptation of LOSA to a Threat and Error Management-based *Clinical Operation Safety Audit* (COSA), and reports a series of 30 observations of trauma team activations in the Emergency Department at an American College of Surgeons accredited level 1 trauma center in the United States of America.

Results of these observations showed discrepancies between work as designed and as executed, as well as other behaviors, associated with increased risk to patients. Analysis of data revealed important areas for targeted improvement based on risk created by the healthcare system during normal clinical operations.

Systematic observation following the COSA protocol can become a vital and essential new tool to assist in improving patient safety in healthcare. The bulk of this thesis considers the criticality of context in work analysis throughout the discussion section. Though concepts of threats and undesired states were easily adaptable to healthcare, error was found to be too narrow a concept. I therefore propose discarding error for a more open and inclusive interpretation of performance: Task Adaptation. We therefore propose to widen our scope and continue to develop Threat Management and Task Adaptation-based COSA throughout the hospital to enhance system performance and improve patient safety.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

INTRODUCTION

This project began as an attempt to answer two questions: How do people work together in health care? Can the evolution of risk be detected during systematic observation of team performance? These arose from the great interest in healthcare for applying efforts perceived to have been useful in other fields such as aviation, especially in terms of teamwork (Frush, Leonard, & Frankel Allan, 2013). The parallels between aviation and healthcare have been well documented elsewhere, but one significant similarity has been the extremely hierarchical relationship between physicians, surgeons, trainees and nurses (Helmreich & Merritt, 1998; Carney, West, Neily, Mills, & Bagian, 2010). The complexity of modern healthcare has produced great specialization in both physician and nursing areas (Jordan, 1985). Nearly every aspect of medical practice has fundamentally changed over the past 75 years, and the diagnostic and therapeutic work is now so widely distributed and interdependent among various team members that the idea of a patient having a single physician dictating care could reasonably be considered fictional (Jauhar, 2014). Perhaps paradoxically, the extraordinary expertise in many individuals – particularly in procedural medicine – is only demonstrable with a large array of other specialists managing increasingly complex tasks (Cassel & Reuben, 2011).

For past thirty years the perception has existed that the inability of medical professionals to function well as a team has degraded care, even before the Institute of Medicine report raised an alarm regarding preventable harm (Leape, 1994; Kohn, Corrigan, & Donaldson, 2000). Social interactions were anecdotally understood to negatively or positively impact performance, but little debate occurred regarding the system within which actors performed (Leape, et al., 2012). The tacit assumption has been that everyone knows what to do and when to do it, and the only dilemma is effectively coordinating the work, and removing a few bad apples (The Joint Commission, 2008). It is in this context that improving teamwork in healthcare is currently considered to be something of a Holy Grail in patient safety.

It is not surprising then that simulation training has come to be viewed as a viable method to decrease harm and to improve task performance (even though very few have suggested such training or other observations be continued after one is in practice) (Gawande, 2011). The only training in which physicians are regularly asked to demonstrate task proficiency is advanced cardiac or trauma life support, and these are the lowest fidelity simulations attainable (ATLS, 2012). Higher fidelity simulation is believed to be a way for us to perfect our skills away from the patient, and decrease error leading to patient harm (Salas, Prince, Baker, & Shrestha, 1995). The advancement of teamwork training and the movement for simulation training have brought areas of task performance and social interactions to the forefront of thinking in healthcare (Frush, Leonard, & Frankel, 2013). Nevertheless, we remain less safe than we would like (Makary & Daniel, 2016).

Our drive for changing rates of patient harm has, to some extent, put the cart before the horse; to alter outcomes we do not like we are attempting to change a process about which we understand little (Perry & Wears, 2012). To understand what effect improved teamwork might have for healthcare, it is important to understand what we already do while working together (Perry, Wears, & McDonald, 2013). We must have some understanding of how teams form and dissolve, how they organize their work, how the system shapes them and impacts their performance (Branlat & Woods, 2010). We must understand the context of their behavior to know how/what/where to make design changes (Conklin, 2016). This is the landscape for the questions that lead this thesis. What lens can we use to analyze how we are working and help us to determine if it is going well or not? Literature review from both the state of teamwork training in healthcare as well as progress through the 1990's in aviation pointed to the development of systematic observation through the Line Operation Safety Audit (LOSA) as a key element in understanding and shaping teamwork performance in the cockpit toward safety (Tullo, 2010; Frush, Maynard, Koeble, & Schwendimann, 2013).

The LOSA was developed to answer precisely the questions raised at the beginning of this paper: what are teams doing, and how are they managing risk (Klinect, 2005)? It was understood at the time that a discrepancy existed between activity demonstrated in simulations or check-line rides versus performance in regular flights, because inexplicable outcomes continued to occur (Klinect, 2005). Activity theory posits that all work consists of goals, tools, rules and roles (Engestrom, 2000). Helmreich and Klinect argued that to understand cockpit work flow in vivo, only a form of invasive monitoring would suffice, and they proposed systematic observation (Helmriech, Klinect, & Wilhelm, 1998). They designed a strict protocol to minimize the Hawthorne effect and move observed behavior as close to "normal" as possible (Klinect, Murray, Merritt, & Helmriech, 2003). These included 10 specific processes, but arguably the most crucial were (1) voluntary participation on the part of the observed, (2) avoiding recording specific identifiers related to any observation, (3) maintenance of confidentiality by observers, and finally, (4) analysis of data only in aggregate form. Each of these represented tradeoffs in ability to make granular corrections of work to gain two specific areas of knowledge about system performance. The first was a more realistic performance by the actor, i.e. as close as possible to normal work. The second was a deeper (and presumably more accurate) understanding of the system. These techniques rapidly led to a theory of cockpit work termed "threat and error management" (TEM) (Merritt & Klinect, 2006).

TEM-based observation proposes that all work analysis must account for two types of variability: the first arising from the environment outside the actors control, and the second from normal human performance. Goals, rules, tools and roles may be essential to organize work for reproducibility, but normal work has the messiness of a chaotic, open system (Hollnagel & Amalberti, 2001). Guided by the work of James Reason, Helmreich et al expected errors to occur and hypothesized that a portion of teams' work was to successfully mitigate these events (Reason, 1998). TEM postulates that successful work is defined by team performance appropriately righting the ship in the face of variability encountered and/or produced by the team (Merritt & Klinect, 2006). This variability can come from forces beyond the control of the actors (e.g. mechanical malfunctions or weather, termed "threats") or be due to the actor's own mistakes, slips or lapses (Reason's taxonomy of "errors") (Reason, 1990). It is important to understand that these factors often alter workflow in significant ways, and furthermore, that teams must use judgment regarding resource allocation to determine the necessity of addressing any specific threat or error (Klinect, 2005). The final element in the TEM model of safe work performance is that occasionally the team runs into a situation of high risk *attributable to their own behavior* – termed an Undesired State – and at this point the team must allocate resources to adjust course or face potential serious harm (Klinect, 2005).

In Klinect's own words, TEM-based LOSA can therefore be considered a "Swiss cheese measuring device", designed to detect how adept the system is at identifying and plugging the holes in Reason's model (as exemplified by the team performance at the sharp end). Klinect is honest about the limits of the Swiss cheese model; in his words "the cheese isn't static – holes are growing and shrinking all the time" (Klinect, personal communication). TEM is simply a conceptualization of how a fluid system reshapes the cheese. From that point of view, TEM could also be considered a method of measuring how effective the system is at pushing work away from Rasmussen's safety boundary through expenditures of dollars and/or effort (again, as exemplified by sharp end performance) (Rasmussen, 1997). In theory, TEM-based LOSA is a measure of both the systems' need for, and success at, flexible adjustment to the environment (Klinect, 2005). Teams regularly prevent harm, and understanding these efforts is important because (a) they are efforts – i.e. they require resources and costs that further alter work flow, and (b) these efforts regularly go unnoticed despite near miss reporting. Often such prevention happens before a near miss could have been considered to occur. Unfortunately, teams' adaptability to the flow of work can also create harm or increase the risk of harm (Reason, 2008). "Teams" here, should not be considered the human error icing on a well-designed system. "Teams" are cognitive functions produced by the system and can in no way be considered separate from it (Salmon, Walker, & Stanton, 2015; Woods & Hollnagel, 2006). Like a chemical equation, risk is not static. It is always being created and mitigated and our desire is simply to understand the reactants and drive the equation in the appropriate direction given specific local conditions (Amalberti, 2013). Attempts to manipulate behavior in such a complex system must understand both where the risk is being created and resolved and this cannot be done irrespective of the local conditions (Braithwaite, Wears, & Hollnagel, 2015). In order to adequately understand the context of team behavior and the emergence of risk and safety, TEM-based systematic observation was explored as a method of evaluating normal work as done in a hospital setting.

This report, therefore, summarizes the findings of the initial Threat and Error Management-based Clinical Operation Safety Audit (TEM-based COSA) done for Trauma Services at a major trauma center in the United States. This project was undertaken as a "proof of concept" study to determine if (1) Threat and Error Management (TEM) based systematic observation methods could be adapted to healthcare and (2) local risk attributable to system function within a specific clinical domain could be determined. If successful, this proactive safety management initiative would accurately describe system behavior that contributes to risk, allowing targeted performance improvement efforts within the system.

The following report is in four parts:

1. Description of the Clinical Observation Safety Audit (COSA) process
2. Brief explanation Threat & Error Management (TEM) and the TEM cycle
3. Review of data
4. Discussion and proposal for adaptation to a new model of work.

This project adheres as closely as possible to the original methods and analysis of the LOSA as a starting point. Terms such as "error" and "risk attributable to team performance" are carried over from the LOSA lexicon. Bear with such phrasing for the time being; the discussion at the end will consider their utility. This discussion and the proposal for adapting to a new model of work serves as the bulk of this thesis.

METHODS

PART I: DESCRIPTION OF CLINICAL OBSERVATION SAFETY AUDITS (COSA)

COSA utilizes systematic observation as a diagnostic test of basic operations following the operational characteristics of the original LOSA design (Klinect, Murray, Merritt, & Helmriech, 2003). The fundamental requirements of a LOSA are as follows:

- Direct observation of the normal operations
- Anonymous and non-punitive data collection
- Voluntary participation
- Trusted and trained observers
- Joint sponsorship of the audits [between staff and organization]
- Establishment of reliable and valid safety-targeted and teamwork-targeted data collection platform
- Trusted data collection site
- Data cleaning roundtables
- Data derived targets for improvement
- Results feedback to audited areas

The underlying imperative in COSA, as in LOSA, is trust. The data collected must be a valid representation of work under normal operational conditions. To understand where risk is created under normal circumstances, one must create conditions of unique safety for the provider, such that the observed trust the process completely. First and foremost, this involves believing they can expose themselves and their thought processes without fear of retribution. Closely related to that is the need for the observed to feel no internal pressure to perform out of the ordinary. Any sense the observed have that they should "turn it on" for the observer must be squelched, and for this reason all observations begin with an introduction by the observer and a request for permission to observe at that specific time. The observed have the right to refuse any observation without question. During the introduction and request, it is stressed to those who are agreeing to be observed that the goal is not to judge their performance but to look at how all aspects of the system respond to each other. If the observed have any internal stressors they feel will prevent themselves from behaving normally, they are encouraged to decline observation.

Observers must be specifically trained in Threat and Error Management observation techniques, and be trusted by both the observed and the institution as being respectful of confidentiality, nonbiased, and competent in their observations. The entire process should be jointly managed, such that neither staff nor management can be perceived as owning either the audit process or the data. Ideally both the staff and management will perceive the COSA as an opportunity for each to learn valuable information they can use to improve system performance rather than any specific individual performance. All data must be collected systematically (achieved through observer training), and recorded in a platform that permits analysis of both threat and error management and teamwork behaviors. (This will be described in more detail below.) Data must be stored in a secure site that does not permit extraction of individual team performance. Even though all data is recorded without specific identifiers, any sense that an individual or team's confidence is betrayed by the COSA process will derail the trust necessary for staff to work as normally as possible in front of the observer.

Observer bias is inherent in systematic observation, so all data must be "cleaned". Following completion of data collection, a Data Cleaning Roundtable (DCRT) is convened, consisting of members of the COSA project team, representatives of hospital Quality and Safety, and representatives of the various observed staff positions (e.g. nursing, physicians, techs, etc.) who were not themselves subject to observation. It is important that these representatives be trusted by both frontline staff and management. Each data point submitted is reviewed aby the data cleaning roundtable to ensure that errors are accurately codified and to reach consensus that data point along with the accompanying narrative is accurately substantiated in the narrative record.

The final two elements necessary for effective COSA are transparency within the organization and utility of the data. COSA data must be presented openly with direct feedback about the data to observed areas (though not specific feedback to observed participants) and targeted areas for improvement. These elements are necessary for the observed to ensure their personal exposure was worth the risk involved. Similarly, the entire undertaking must provide actionable data for the institution to perceive adequate return on investment. This initial COSA adhered closely to those operational requirements as outlined above.

OBSERVER TRAINING Observer training is a two-step process. The first step is a four-hour course in teamwork skills. This is followed by a two-day classroom session teaching concepts of:

- Threat and Error Management Cycle
- Effective note taking
- Narrative writing
- Classification of Threats, Errors and Undesired States
- Observation conduct – especially including voluntary nature of the observations and the need for observer to maintain an anodyne presence during observation

This endeavor is the first of its kind, and there is currently no standard for TEM-based systematic observation training in healthcare. The author sought guidance from Dr. Klinect, who generously permitted the author and an anesthesiologist colleague to attend training intended for pilot observers as guests of the LOSA Collaborative and Hawaiian Airlines. The training uses vignettes to standardize observers to identify teamwork skills and view each observation as a series of team threat and error management events. Initial trial observations were done in the operating suite by both of the observers who had attended the pilot training. Five of these observations were done as a test by simultaneous observation of operative cases by the two observers as well as an operating room nurse. Narratives and coding of TEM by the teams were done separately by these three observers, and these were then compared to assess our perceived ability to adequately capture the nature of events both socially and clinically. The only discrepancies noted in these comparisons were domain specific technical issues within our respective realms (anesthesia, nursing and surgery). Based on these comparisons, we believed the systematic observation methods involved in LOSA to be both adaptable to healthcare, and worth the larger evaluation described herein. None of those trial observations are included in this analysis, as they served as a learning experience for the author and the hospital. The data collected for this study was collected by a single observer (the author), and subject to review by the usual LOSA protocol for data cleaning roundtable, further described below.

OBSERVATION PROTOCOL and DATA COLLECTION Each observation begins with the observer formally requesting permission to observe the providers, and reminding them that (1) it is the overall system performance that is being evaluated rather than the individual provider, and (2) no identifying data will be recorded with respect to any individual observed.

Duration of the observation for the Trauma Activations during this pilot study involved only the period between trauma alert page from the operator until the patient's work-up in the trauma bay was complete and the patient was moved to the next stage (i.e. to CT scan, the operating room, etc.). The start and end points for observations are inherently somewhat arbitrary, and the most important point is that they are adequately established to capture useful data for the clinical area/work being evaluated.

Each observation concludes with the observer thanking the observed. Our protocol adds two questions to the observed at this time:

1. What do you perceive to be the greatest safety risk in your work environment? (This question is purposely left open to be answered with respect to staff or patients.)
2. Do you feel the presence of the observer altered either your or anyone else's behavior during the observation, and if so, how?

This data is recorded with each observation for analysis in aggregate.

Observations record behavior of the team in note format, and immediately following the observation, a detailed narrative is constructed to provide context for all team behaviors. It is extremely important that this is done while the observation is fresh in the mind of the observer for adequate contextual understanding of the event. This narrative provides context for the threat and error management strategies employed. The data is logged into a password protected computer. The goal of the observer is to collect descriptive data of team performance with respect to the TEM cycle (as described in part II below) and teamwork performance markers (as described in part III below). Approximately 80% of a high-quality narrative will be descriptions of human performance with respect to the TEM cycle, and the remaining 20%

should be narrative glue to provide context for both team decision making and rationale for observer coding. (LOSA Collaborative observer training, Hawaiian Airlines, 2014)

In addition to team management strategies, milestone events within the trauma activation were recorded for all observations. These included timing of specific events such as Trauma Attending arrival, patient entry, first and successive vital signs, blood gas, labs, abdominal ultrasound examination, and chest x-ray, were recorded as able based on other events occurring during the activation. Inclusion criteria for observations included trauma level I or II and availability of observer to observe case in its entirety. Trauma activations were considered as having three separate phases based on American College of Surgeons' Advanced Trauma Life Support (ATLS) guidelines: Prearrival, Patient Entry, and Survey/Resuscitation. The Prearrival phase began as soon as the trauma alert went out through the paging system, and ended with patient entry into the room. Patient Entry began with Emergency Medical Services (EMS) bringing patient into the room and ended with patient being moved from EMS stretcher to the trauma bay stretcher. The ATLS phases of Primary and Secondary surveys and ongoing resuscitation could rarely be clearly separated, so were considered together as the Survey/Resuscitation phase, starting with patient positioning on the trauma bay stretcher. Detailed analysis of the team behaviors documented in the narrative followed the Threat and Error Management Cycle.

ETHICAL CONSIDERATIONS The project was undertaken as a Performance Improvement Project within the Trauma Service, under a waiver from the Institutional Review Board. As a quality improvement project, patient consent is deemed unnecessary.

No patient identifiers were recorded. Patient information was only recorded with respect to age, history as reported by the emergency medical transport team, and vital signs during the course of the trauma activation. No staff identifiers were recorded. All staff members were described by title only, and no sex pronouns were used in the narrative to further protect the observed. In all cases, consent was obtained from the staff to be observed prior to the observation. Time of day and day of week were recorded, but not date of observation. It should be noted that the observer is obligated to intervene if they see the potential for imminent patient harm.

TEAMWORK PERFORMANCE MARKERS The narrative was deconstructed to document all instances of teamwork skills behaviors (as shown in table 1) and Teamwork Performance Markers (TPM) were scored on a Likert scale in the domains of Overall Climate, Planning, Execution and Review/Modify. Each of these domains has been considered to contain teamwork skills critical for threat, error &/or undesired state management. (see Table 1). Each team was thus evaluated separately for Teamwork Performance Markers and Threat and Error Management.

Table 1: Teamwork Performance Markers

OVERALL CLIMATE	*PLANNING*	*EXECUTION*	*REVIEW/MODIFY*
Leadership	SOP Briefings	Monitor/ Crosscheck	Inquiry
Communication environment	Contingency management	Workload management	Evaluation of plans
Overall performance	Plans stated		Assertiveness

DATA CLEANING Each Threat, Error and Undesired State was reviewed through the entire management cycle during data cleaning round table discussions to arrive at consensus with respect to validity and accuracy. Representatives of each major trauma team role were present for these discussions, including trauma surgeons, emergency department physicians, nursing, and surgical residents. Additional representatives from hospital patient safety leadership, and TeamSTEPPS® master trainers were also present at each discussion. During this process no data points were discarded, but several threats and errors were added based on data from the narratives.

STATISTICAL ANALYSIS Microsoft Excel was used to aggregate the thirty observations and to perform graphical and statistical analysis of data. Prevalence of threats and errors were tabulated per the following qualitative variables: case, phase, category, outcomes and response.

Comparisons between categorical variables of team response versus outcome were tested by use of contingency tables to generate a management index. The "management index" is the percentage of a threat or error that are linked to downstream error. Threat and error responses were grouped as 'aware' versus 'unaware'. An 'aware' response is defined as a threat response of 'anticipated' or an error response of 'detected and acted upon' or 'ignored'. An 'unaware' response is defined as a threat response of 'pop-up threat' or 'unanticipated', or an error response of 'undetected' or 'detected by external'. Aware versus unaware responses were divided by linked outcomes for determination of management index. 'Linked' outcome is defined as a threat outcome of 'linked to team error' or an error outcome of 'additional team error' and 'undesired state'. Chi-square test was used to determine significance between relationship of response awareness and linked outcome.

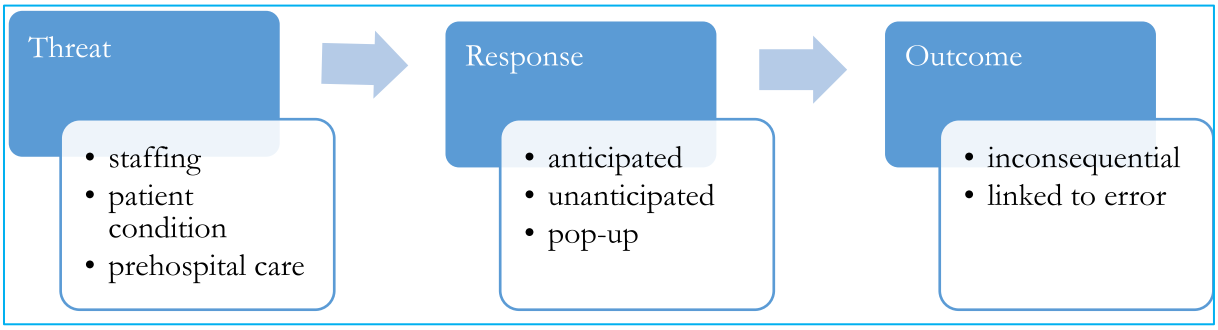
PART II: THREAT & ERROR MANAGEMENT AND THE TEM CYCLE

The primary analysis considers threat and error management (TEM) strategies performed by the team. TEM is founded on the premise that nearly all team activities involve threats the team must manage and small deviations from expected behavior (errors) that influence the team's or patient's exposure to risk. Increased exposure to risk secondary to team error is considered an "Undesired State". Threats, Errors, and Undesired States are defined as follows:

Term	Definition
Threat	An event that occurs outside the influence of the healthcare team, increases the operational complexity of the care, and requires team attention and management if safety margins are to be maintained.
Error	Action or inaction that leads to a deviation from team or organizational intentions or expectations.
Undesired State	A position or condition that clearly reduces safety margins *and is the result of actions by the healthcare team*

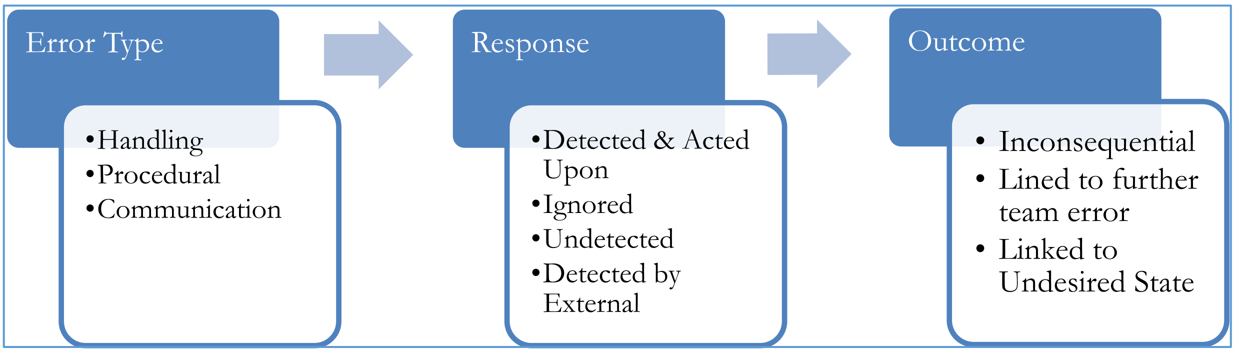
TEM-based COSA evaluates what teams do or do not do to manage these situations as they are confronted by them, and describes the conditions that give rise to increased risk attributable to team behavior. Each threat is broken down as follows: description of the threat, description of steps taken to manage (if any); phase of occurrence; team member (by role) who identified the threat; team response; and outcome. "Team response" to threats is considered in terms of team preparation. For example, team response to each threat could be considered "Anticipated" – team considered the threat and took steps to manage it, "Unanticipated" – team could have planned for the threat but did not, or "Pop-up Threat" – team had no opportunity to consider the threat prior its presentation. Each threat was considered to have a binary Outcome: either "Linked to Team Error", or "Inconsequential".

The threat → response → outcome cycle, with examples, is shown below:



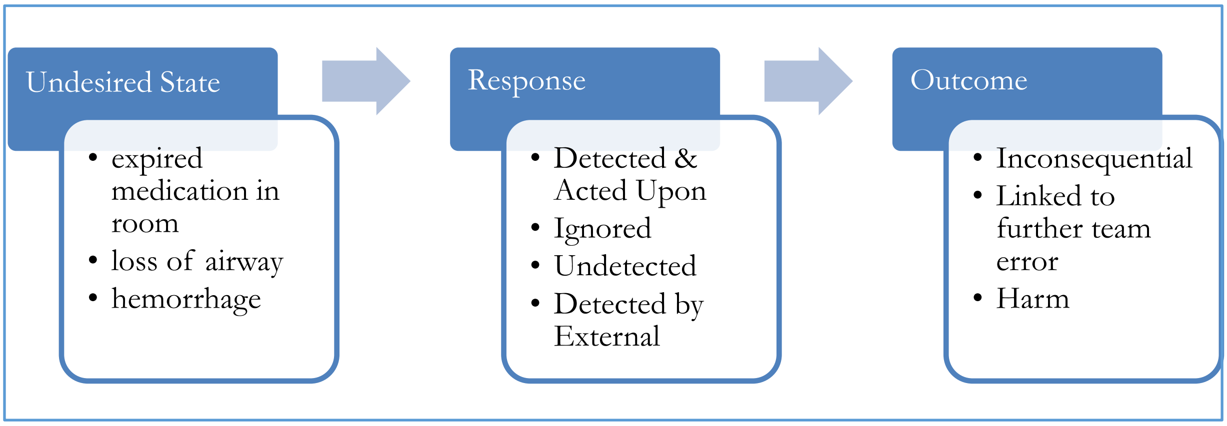
Errors were defined as deviations from behavior expected by either the team member or the organization. Errors were broadly categorized as Handling, Procedural or Communication types, and further broken down into subcategories as appropriate. Handling errors were considered to fall under the domains of Technical Proficiency (e.g. intravenous line placement) or Judgment (e.g. decision to transfuse). Procedural errors are equivalent to "Standard Operating Procedures", such as use of a Pre-arrival Checklist, maintaining sterile precautions, or crosschecking a bed being locked prior to transfer. Communication errors are categorized as occurring within the team (Team-Team), between the team and the outside world (Team-External) or between the Team and the Patient.

Separate descriptions of each error managed by the team included: description of error, description of management, phase of the occurrence, team member (by title) responsible for any error, team member (by title) who identified error, and outcome. Team response to error considered two factors: their awareness of the error and actions taken. Teams that detect an error may consider the error important to fix and act upon it in some way, or they may choose to ignore it. On the other hand, an error may go entirely unnoticed by teams. At other times the potential for harm may be so high that the observer or another outside entity may be the first to notice the error and bring it to the teams' attention. All Error outcomes were either considered Inconsequential, i.e. they led to no further error or undesired state, or Linked to another event (either subsequent error or undesired state). The Error → Response → Outcome cycle is shown schematically below:



Within the scope of this project an "Undesired State" is defined as a situation where harm (to patient or staff) has been made significantly more likely *as a result of team error*. These are generally categorized in terms of patient physiology as relating to airway, breathing, circulatory, metabolic or neurologic compromises of some type. In addition, they can include "Configuration State", which is a state of physical positioning of the patient that is known or presumed to increase risk. Examples of configuration states are: moving a patient without the bed locked, failing to stabilize the cervical spine in a trauma patient, having the wrong site prepped prior to a procedure, or expired or incorrect instruments/medication present in a room with a patient.

Undesired State (US) management and outcomes are considered in a similar fashion to errors. Each US is described and then analyzed for team management, response (with options identical to Error responses) and outcome. The possible outcomes for an Undesired State are Inconsequential, Linked to Team Error or Harm (to patient or staff). Examples of Undesired States and the Undesired States Cycle are shown below:



Using the methods above, TEM-based COSA captures the entire management process, such that each Threat, Error and Undesired State is described relative to the team response and the outcome.

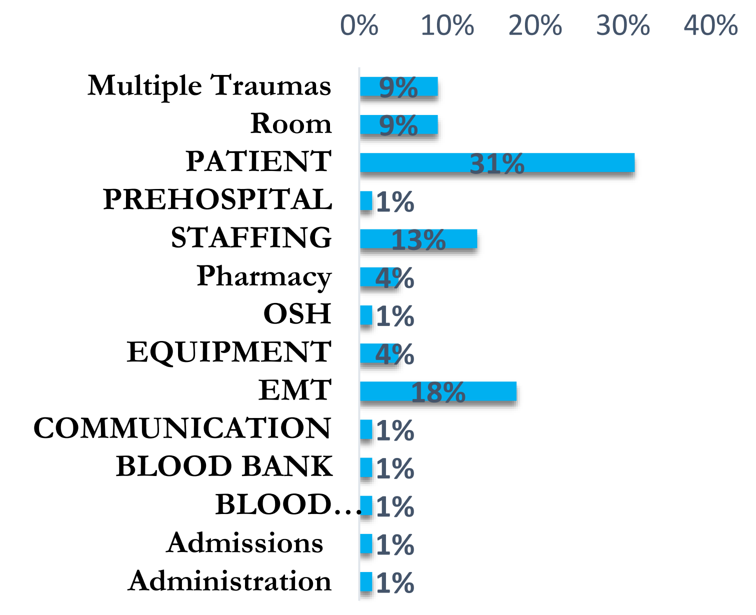
Through this process, team behavior is given context as threat and error management behaviors are understood. Each observation is considered as a series of Threat → Response → Outcome, Error → Response → Outcome, and occasionally Undesired State → Response → Outcome cycles. Each observation therefore unfolds as though the team is traversing a maze, navigating and managing obstacles, some of which are their own creation.

RESULTS

THREATS

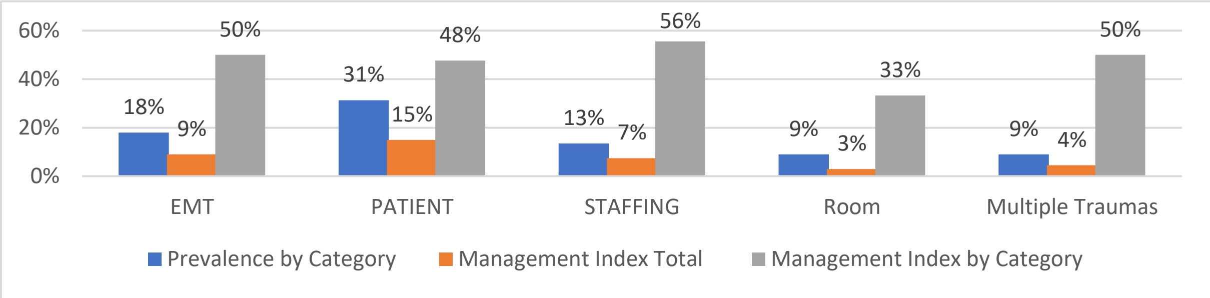
Every team had to manage complexity outside their control generated by another portion of the system (from prehospital care to our own hospital). Average threats per team was 2.4, with a range from 1-7. 31% of patients could be considered to have conditions that were their own threats, for which ATLS algorithms are designed. Total threat prevalence by type is listed in Table 2.

TABLE 2: Threat Prevalence



Sample size limits interpretation of less frequent types, but for threats occurring in more than 5% of the cases a Management Index could be calculated based on the threat exposure and the likelihood of linkage to a team error. This is shown in table 3.

TABLE 3: Threat Management



As an example, 18% of teams had a threat associated with Emergency Medical Technician (EMT) management of the patient, often something as simple as failing to obtain spinal stabilization. 9% of all teams had linkage of a threat posed by EMT to a team error, leading to a 50% rate of mismanagement of threats associated with EMTs. Generally speaking, with respect to threat management, our system performed as desired about 50% of the time (i.e. teams managed threats without linkage to a downstream "error" approximately 50% of the time). It is nearly impossible to interpret this raw number without an understanding of context. One simply cannot say whether such a high rate represents appropriate ability to adapt performance to local situation, or whether system constraints on performance are too loose or even unrealistically tight. We can, however, draw some conclusions regarding the utility of threat preparation when considering linkage rates among various types of threat response. The analysis of threat management incorporates planning, as threats are considered to be of two basic types: pop-up, or known. Known threats can be anticipated or unanticipated (i.e. the team could plan for managing the threat, or they could simply decide to improvise should the threat present itself[1]). Our evidence supports the benefit of planning, as threat management was significantly improved when teams considered the threat in advance: When considered in terms of threats that were anticipated, i.e. planned for through open discussion by the team, linkage to error dropped from 53% to 21%. This was statistically significant, with a p value of 0.01.

[1] COSA makes no judgements about whether threat anticipation is good or bad. We are concerned with what pattern of downstream actions is associated with threat anticipation v. non-anticipation.

ERRORS

Errors were considered to be one of three types as described by Klinect (Klinect, 2005). Procedural errors are failure to follow defined protocol or policy. Communication errors may be failure to use standard communication techniques, such as read-back of information, or as obvious misunderstandings during communication. Handling errors are largely technical in nature, but could also include judgment. As shown in table 4, Procedural errors were far more common in this series. Management indices were equivalent between error types, but procedural types accounted for significantly more Undesired States than either Handling or Communication errors. One out every five errors was linked to an Undesired State, and over 50% all Undesired States were associated with Procedural errors. Most of these were associated with configuration states encountered during patient entry.

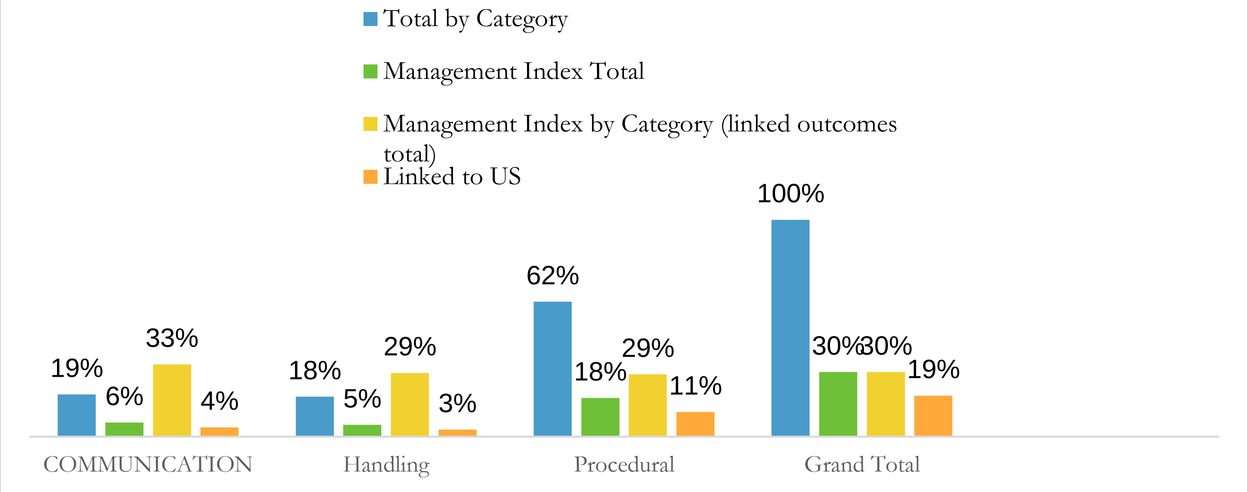


Table 4: Errors with Linked Outcomes

HANDLING ERRORS Handling errors showed no patterns whatsoever, and appeared to be the result of random slips, lapses or mistakes as expected. Only four of the linked Handling errors could be considered technical or skill based.

It may be notable that the Undesired States appeared more associated with judgment errors than technical proficiency, though the sample size is small and the categorization is too difficult to say with any certainty at this point. It is also notable that many of the events we consider "errors" are so rapidly resolved during normal workflow they can appear and disappear as quickly as keystroke mistakes that are 'autocorrected' by the computer on which this is typed. This is a critical value of such an observational process, and will be revisited in the discussion.

COMMUNICATION ERRORS Communication errors occurred almost entirely within the team. Five out of seven of the linked communication errors led directly to Undesired States.

PROCEDURAL ERRORS Procedural errors, like Handling errors, were wide ranging, but unlike Handling errors, they occurred with a frequency that did permit some patterns to emerge. Overwhelmingly, the procedural errors were related to either:

I. the communication environment in the room (through a failure to maintain sterile communications, failure to callout information or give/receive report),
II. failure to crosscheck leading to configuration state
III. failure to use checklist

The checklist was *never* used as intended – i.e. with full team attention, completed, and in sterile fashion. In addition, for two thirds of the study it was inaccurate because of policy changes within the trauma service that had not been refleted in the printed checklist in the room. The most complete checklist usage witnessed was when the entire list was run through by the team leader (almost always the resident prior to trauma attending arrival), but never with full attention from the team. Because this was the cultural norm, errors were only coded when deviations from that standard occurred.

More than half of the linkages for procedural errors led directly to Undesired State (20/33). These consisted of: 17 crosscheck errors leading to configuration states or loss of spinal precautions, 1 nonsterile handoff, 1 failure to callout, and 1 failure to complete the checklist.

PHASE EFFECTS ON THREAT AND ERROR MANAGEMENT

All threats, errors and undesired states were coded for phase of activation, considered to be either Prearrival, Patient Entry or Primary/Secondary Survey. Rates of "Linked to US" derived from: Linked to US/total error in phase.

Table 5: Phase effects of Threat & Error Management

PHASE	THREAT PREVALENCE	MI	ERROR PREVALENCE	MI	LINKED TO US
Prearrival	42%	39%	16%	28%	10%
Patient Entry	19%	46%	23%	48%	38%
Primary/Secondary	39%	42%	62%	24%	14%

Notable findings include:

- Management index was worst for both Threats and Errors during Patient Entry phase.
- Errors occurring during the Patient Entry stage had an extremely high likelihood of leading to US (approximately twice the rate of errors at other phases).

Undesired State linkage by phase showed other characteristics as well.

- All 3 Prearrival errors leading to US created the same type of US, and were caused by the same threat. Despite this similarity, they differed with respect to error type (one each for Handling, Communication & Procedural) as well as incident response (one undetected, one ignored and one detected and acted upon).
- All Patient Entry errors leading to US were Procedural Errors leading to Configuration State (described below).
- 80% of mismanaged threats occurring during Patient Entry led to a subsequent Undesired State, and these were just as likely to lead to a physiologic Undesired States (e.g. diagnostic error/delay, hemodynamic instability).

Primary/Secondary Survey errors leading to Undesired States were relatively evenly split between error types. Again, most of these (57%) were conditions other than a Configuration State.

UNDESIRED STATES

Undesired States occurred 35 times in our series, and in 70% of the trauma activations. Level I trauma activations occurred 14 times in the series, with 12 (86%) presenting direct from the field and 2 occurring as transfers. The level I activations accounted for 57% (20/35) of the Undesired States, and nearly half of the observed US (45%) in level I activations were *not* considered Configuration States. In contrast, only 20% of the undesired states in the level II activations were not related to Configuration States. Configuration States were nearly always the result of procedural errors, and occasionally (two instances) related to communication errors. They were also almost entirely related to one of the three issues:

I. failure to secure bed prior to patient transfer
II. failure to stabilize cervical spine prior to transfer
III. failure to protect patient with side rails being up when leaving patient unattended

For Undesired States not categorized as Configuration States, error types were very different. In these cases, 62% (8/13) were related to Team-Team Communication errors, 38% were related to Handling errors (5/13 - all judgment errors), and only one occurred subsequent to a Procedural error (nonsterile checklist). Half of the communication errors were failure to acknowledge communication, and the others were some type of failure to offer information (failure to assert, failure to callout finding, offering unclear directions). All the judgment errors were errors of omission, commonly considered to be related to poor situational awareness (not attending to novice staff, not reconsidering medication dosing given patient's history, and not attending to the need for sterility during central line placement).

TEAMWORK PERFORMANCE

All teams were also evaluated for Teamwork Performance Markers (TPM) on a Likert scale in the domains shown in Table 1 (page 10).

Analysis of Teamwork Performance Markers showed our teams were normally distributed with respect to demonstration of teamwork skills. We sought to determine what differences existed with respect to TPM between the teams a standard deviation above and below the mean for risk mitigation. To do so, we considered both the teams' propensity for error as well as their propensity for poor management of threats and errors and established a Management Ratio as shown below: Linked Events/Total Error = Management Ratio. A lower Management Ratio would correspond to a teams' improved ability to autocorrect for error.

Figure 1: Correlation of TEM performance and Teamwork Performance Markers

We then compared the teams a standard deviation above and below the mean Management Ratio in terms of their Teamwork Performance Markers (TPM). Figure 1 shows "Good TEM" performers in blue (those with a better Management Ratio) and "Poor TEM performers in orange. The Y-axis is the average Likert scale score for Teamwork Performance Markers. There was a strong trend toward a significant difference in domains of Planning, Review/Modify and Overall Climate for teams who were better TEM performers (p=0.1). Given the very small numbers involved it is not surprising that we failed to see statistically significant differences between groups. Despite that, every one of the "Poor TEM" teams had at least one physiologic Undesired State (some had multiple), while only one "Good TEM" team had a single Undesired State. This difference between the two groups was highly significant (p=0.001)

DISCUSSION

Although the chronologic flow of teamwork as considered by TEM is threat → response → outcome, error → response → outcome, *potentially* followed by creation of an Undesired State, I will begin with US in this discussion, because they represent at least a "near miss", and often a "no-harm" event.[2] (The Joint Commission, 2016) Rather than consider them as the fault of the team, *we should consider undesired states as risk resulting from normal system function*.

[2] Remember that Undesired States exist in two basic categories: physiologic abnormalities that must be corrected or configuration states known to produce high risk of harm.

Irrespective of the COSA, with respect to outcomes our trauma service regularly has an observed to expected mortality ratio of less than one, and this was no different during the period of these observations (Trauma Registry data). From an outcome point of view, therefore, these teams function particularly well. Furthermore, the COSA process observed no harm occur to any patients or staff. Systematic observation of system function nevertheless revealed a high prevalence of undesired states. These are meaningful to us in the same way that near miss reporting is meaningful: they represent opportunities for the system to learn and improve. One critical value of COSA is the ability to see work flow with much more granularity and decipher patterns within the system (Klinect, 2005). For example: during the period of our evaluations, we had a 6% rate of anonymous incident reporting during the first 24 hours of a patient's stay on the trauma service (Hospital Quality Data). Our observations, however, showed Undesired States occurred in 70% of the trauma activations and 35 times in our series. This represented risk created during, on average, the patient's first 22 minutes in the hospital. What risk did this represent?

First consider configuration states, as these were both the most prevalent and, also, because they can be tricky to understand. In our series they were overwhelmingly one of three types:

I. Transferring a patient without cervical spine stabilization
II. Leaving patient unattended with side rails down
III. Transferring a patient from stretcher to bed without securing either of them

Each of these represents a known risk which, though very small, is also very real. We have protocols in place to make sure that none of these events occurs, and yet one or all of them occurred more often than not. Each of these was also associated with a unique pattern of system function, as described below.

Maintaining cervical stabilization receives the highest degree of attention by the system of these three. The potential for cervical spine injury is discussed often, in every trauma patient. Teams jockey over appropriate removal of the prophylactic cervical spine collar placed to protect each patient on a daily basis throughout the hospital. The observed teams commented on this in one way or another for each of the three cases involved where a collar was not immediately applied. In the first case the team leader assumed (correctly) that a collar would not have been placed in the field, based on the prehospital report. From a TEM standpoint, this would be considered an anticipated threat, because the team planned for an expected additional work due to the high potential for a cervical injury, the expectation that patient would not have had collar placed prior to arrival, and the desire to follow protocol. Nevertheless, on arrival the team was consumed with attention to the report from the emergency transport team and for beginning to care for the patient, so no collar was placed until after the patient was moved. In the second case, the absence of the collar was noticed only after transfer of the patient to the bed. Prehospital providers had attempted to place the collar in the field, but they could not because the neck was "too short". They improvised with a towel taped to the backboard because of their perception of a very high potential for cervical injury in that patient.[3] In the third case, the patient was transferred from another hospital with a known spine injury. This patient arrived without a collar and a report that the spine had been cleared at the outside hospital. Since he was being transferred for a spine injury he had a higher than usual chance of an associated cervical injury, and it is a protocol to not accept outside assessment of the spine because transferring centers often perform what our radiologists consider to be substandard imaging (ATLS, 2012). Nevertheless, in this case no-one questioned the outside reading or the decision to leave a collar off the patient.

In each of these cases there was clear deviation from standard operating procedures, and there was at least some increased degree of risk. This is made more concerning when considered against the backdrop of the remaining 22 cases where a cervical collar was indicated by protocol[4], all of whom either presented with collars in place or had cervical collars placed immediately. Most notably, in the four of those cases where a patient presented without a collar and a collar was placed by the trauma team prior to patient transfer to the trauma bed from the EMS stretcher, protocol was followed even though the acuity of the patient was very low and *there was virtually no specific risk given the patient's injury and presentation*. In the three cases where collar placement was indicated because of mechanism but it was not placed, however, the patient was in a more precarious state, though in none of them was the patient exhibiting any signs of acute physiologic instability. In these cases, the staff – even having planned for the fact that a collar might not have been placed by the prehospital providers – focused on other issues they immediately felt to be important. We can only hypothesize why this is based on the observations, but we can clearly say at this point that the system was aware of the need for a collar and produced it as necessary 88% of the time, in a

[3] This unfortunately falls outside the scope of our evaluation, but their adaptation is fundamental to the argument for disabusing ourselves of the concept of error. The EMTs practiced what medicine has been since Hippocrates: They understood the potential problem and worked to solve it with the tools they had.

[4] Five of the cases were penetrating traumas where a collar was not indicated.

situation where missing an occult injury can be catastrophic and would certainly be considered a "never event".

The second most common configuration state occurred during a moment where each patient gets a chest x-ray approximately 10 minutes into the resuscitation. The entire staff leaves the trauma bay during the x-ray in order to avoid unnecessary exposure to radiation, and the patient is left alone on the stretcher. Some x-ray technicians raised one or both side rails on the stretcher and some did not. There is no hard protocol for this moment within the trauma resuscitation, though there is a basic principle that patients who should not be out of bed, and certainly patients who cannot follow instructions, are not left unattended with side rails down. Despite the patient being left alone for only a matter of seconds, there were at least two unrelated instances where patients who were believed to be sedated suddenly sat up. Both of those times happened to be when staff was in the room during other portions of the trauma team's work, but it would seem to be a simple stochastic event at some point leading to a patient falling out of the stretcher when the team has walked away for the x-ray. During observations, however, the odds that the railing would be raised approached that of a coin toss.

The third, and by far the most common configuration state, was transferring the patient from the EMS stretcher to the trauma bay stretcher without pausing to make sure the stretchers were stabilized. Unlike the previous two, observation of this team behavior yielded a consistent pattern of failing to check the stretchers were stabilized. Patients do fall to the ground during such transfers (though extremely rarely) and severe injuries and even fatalities have been reported. It is expected practice, therefore, that the team checks the two surfaces (stretchers and/or beds) to be locked prior to transfer. From an ergonomics point of view, this is an extremely weak barrier defense in that it relies on human memory for occurrence (Reason, 1990; Kahneman, 2011). Nevertheless, there is currently no stretcher technology that produces automatic locking and requires human involvement for initiation of movement, so the action relies on the collective memory of the team. The odds of the system successfully remembering this based on my observations were nearly the inverse of the likelihood the team would place a cervical collar, i.e. approximately 15%.

THE CRITICAL IMPORTANCE OF CONTEXT IN EVALUATING WORK

It is important to understand that these are extremely different types of events, appearing to represent extremely different levels of system awareness, system activity, and levels of system responsiveness. The consistent theme of configuration states was deviation from "standard procedure". From a work flow standpoint, there might be reasons for deviation from procedural recommendations in any given instance, and a key benefit to systematic observation is the ability to provide a context for better understanding of why procedures are or are not followed (Klinect, 2005). In each of the three instances when cervical collars were not placed, for example, the system was moving forward with a focus on an issue and unconsciously deprioritizing protocol. This is best exemplified by the desire of one team to prepare for an expected absence of a collar in an extremely high-risk patient and still not placing it. In this instance the team leader acknowledged that they simply forgot about it once the patient arrived and remembered only when the patient was being moved[5]. From the standpoint of the observer, this appears to arise from the nature of complexity within the trauma bay environment. Protocols were reliably followed when they were part of routine practice (e.g. placing a collar) in situations of low acuity. As soon as cognitive stress occurred because of the sense that the patient might have a "real" injury, the team's ability to focus on the cervical collar was low even with a high degree of suspicion that a dangerous injury could be present. To understand this, one must consider the broader context of the trauma bay.

Trauma resuscitation is an unpredictable, high stress environment. Walking into a trauma bay to prepare for a resuscitation is accepting the fact that you may soon be watching someone dying in front of you. Many of those deaths are unavoidable, but it is still a person – son, daughter, mother, or father – who is no more. Furthermore, every team member knows that some patients who come in will require rapid decision making and procedural expertise in the first 20 to 60 minutes or they may die when they potentially could have been saved (Scott, Hirschinger, Cox, Brandt, & Hall, 2009; Ertel & Kellam, 2015). This makes the presentation of a patient with a "real" injury automatically more stressful and demand different cognitive energies (Woods,1988). In this context cervical spine injury can lose a great deal of importance because patients simply won't die of an occult cervical cord injury while they very well may die of airway, breathing or circulatory abnormalities that are correctable through adequate team action. Unfortunately, those same patients may also be harmed by the very system designed to save them if some protocols are not followed. While this may not be news to those involved in patient safety, we are left with the undeniable fact that the cognitive demands of that stressful environment reliably preclude execution of nonemergent standard

[5] This is apparent in the recorded conversation in the room at the time of the observation.

protocols (at least as observed in this study). As if that wasn't a difficult enough situation, even though witnessing someone severely injured or dying is a real possibility, 70% of the traumas in this series had no injuries whatsoever and the team understood this within seconds of their arrival. In all of these relatively more benign cases a collar was placed even when the team felt it was unnecessary from a medical standpoint. They nevertheless dutifully followed procedure – or, rather, they had the cognitive space to follow procedure. The primary difference in the cases without collar placement seemed to be focus on a potential real problem. Acute concern for a known problem appeared to prevent attention from being given to protocol. This cognitive dilemma has been well-described (Kahneman, 2011).

At this point, the utility of the COSA is to turn the information back over to the system for resolution. The data is transparent and able to be reviewed by both sharp end providers and those who manage the system, and it is up to them to work out solutions with guidance from human factors/system safety experts. The above description is undertaken only to emphasize the benefits of systematic observation of system function at the granular level of work.

THE HIDDEN COST OF RESILIENT WORK

Though Undesired State data alone might seem to be adequate rationale for placing an observer to identify risk, we are concerned with how the system tends to behave and if we are to alter the system we must have a sense of both what it currently yields and how it does so. Ultimately, we are concerned with identifying learning opportunities. The remainder of the TEM cycle can be very instructive about the context within which such risk arose. For example, threat analysis is concerned with creating a "threat profile" (Klinect, Murray, Merritt, & Helmriech, 2003). What types of complexity does the system have to commonly manage? How good are teams at managing threats? What threats, or combinations of threats, appear to overwhelm team function? A system must have answers to these types of questions if it is to appropriate limited resources effectively. Threat analysis helps a great deal, possibly far more so in medicine than in the cockpit, because patient illness/injury is considered a threat, and therefore our work in the hospital could fundamentaly be considered the management of threats. Consider for a moment that only a third (31%) of the patients entering the trauma bay have real injuries to be managed. All other threats we observed need to be analyzed in terms of that number, because low staff, inadequate supplies, multiple traumas and deviations from protocol by EMTs are extremely unlikely to generate risk in a healthy patient[6]. Our data shows the system to be fundamentally handicapped in the ability to care for the people who need their lives saved, and yet the outcome data contradicts this (remember our O/E mortality ratio is regularly less than 1). This likely contributes, albeit paradoxically, to the healthcare system's inability to move the bar on patient safety. We rely on the resilience of our sharp end to keep things going with consistently inadequate resources, and they are so good at it they defeat their own arguments for changing resource allocation to create space for improvement. Process evaluations such as COSA are essential to bring such issues to management's awareness.

As a method of deciphering path dependency, Klinect analyzes each event, whether Threat, Error and Undesired State, in terms of a management index (Klinect, 2005). This is simply the percentage of time that any one of those events (T, E or US) is associated with another downstream event (E or US). Though satisfactory in the most superficial sense, the management index is only useful for identifying areas of high deviation from expected performance; it makes no value judgment regarding either system design or performance. Any value judgement must be either brought to the table by the analyst or constructed based on the context. When considered in conjunction with teamwork behaviors, however, we can begin to see another pattern emerge. A clear correlation begins to develop between certain teamwork behaviors as tools for effective TEM. In this series teams that anticipated through planning were much better at recognizing and responding to unfolding events and risk was seen to drop precipitously. Effective TEM appears, therefore, to be: (1) facilitated by measurable teamwork behaviors, and (2) decrease risk of harm. Teamwork behaviors can be considered the human tools for *resilient* Work as Done. Where the communication in the room was free-flowing, and when the leader suggested one or two ideas for how to proceed at the beginning, two differences were apparent: the team members showed a strong trend for expressing their own ideas and concerns during the progress of the resuscitation, and the teams had a marked decrease in physiologic undesired states. Simply put, those teams were able to solve problems that

[6] Inadequate room supplies, for example, all revolved around not having some material needed to treat one of the critically ill patients. Since the team only looked for that equipment when they had strong suspicion for a problem or actually needed it (in the 31% of cases where the patient was ill or reported to be ill), it only appears to be a problem 9% of the time. It is more meaningful to say that 29% of the time the room was not ready. This is both concerning and also speaks to how extremely good the teams are at working around inadequate resources.

arose without incurring as much risk. But remember what problems they were solving: inadequate supplies, inadequate staff and, of course, some of their own creative deviations from work design.

Before turning to a consideration of errors within TEM, consider that the Line Observation Safety Audit was developed in the mid-1990s to gain an understanding of the discrepancy between work as imagined and work as done in the cockpit. Within this construct, error could easily be defined as "behavior differing from work as imagined". Threat and Error Management (TEM) was subsequently developed to organize analysis of task performance by teams (Klinect, 2005). TEM, despite such an arbitrary definition of error, is able to function not unlike Newtonian theories of gravity and light were able to be useful even if fundamentally untrue. Like Newton's work, TEM functions as a pretty good approximation of what is going on, and, perhaps more insidiously, feels as though it represents the world as we experience it (Rovelli, 2016). Just as with physics, however, we must ask ourselves if what feels true is the best representation of the reality.

CONSIDERING THE UTILITY OR ERROR

Modern neuroscience and psychology postulates that our brains are prediction machines, designed to guess what will happen next. This permits efficient organization of our internal system to manage ourselves in the world (Seligman, Railton, Baumeister, & Sripada, 2016; Barrett, 2017). Predictive organization is an ongoing and background process occurring beyond the bounds of consciousness, and we are incompletely aware of it (Kahneman, 2011). This process is undergoing constant autocorrection based on moment to moment experience of the world and our internal processes (Barrett, 2017). In this model, we explore our work space much as Rasmussen described, modulating effort for probability of success and at the same time creating a new work space with our own behavior (Rasmussen, 1988). This workspace is the result of our muddling through instantaneously appearing data from both external components in the system and the internal workings of the participant actor (Kahneman, 2011). Some of that guesswork will be inaccurate, increase the potential for inappropriate behavior, and possibly introduce risk of system failure. Thus, we appear to create a probabilistic path dependency through our behaviors as a function of the instantaneously present system state. Some have called the actor's point of view of the system situational awareness, and promoted maintaining an accurate understanding of it (Endsley, 2015). It should be easy to see the idea of *maintaining* situational awareness to be meaningless, however, for it is always being constructed and deconstructed (Dekker, 2015).

Klinect and Helmreich propose that in evaluating system performance we must therefore be concerned with the relatively successful autocorrection of system components, both individually and jointly. In that respect TEM is really the outward manifestation of work that is going on internally in all of us. If guesswork is the basis for our representation of the world, we should expect both inaccuracies and correction in order for successful function. This is how threat management has value as shown above: anticipation of possible problems appears to facilitate better recognition of problem states and their correction.

In the context of real time work, however, the concept of error becomes so diminished and arbitrary as to become meaningless (Hollnagel & Amalberti, 2001). Consider a concrete example from one of the observations as recorded.

> Trauma pager went off at 1816 as follows: Level 1, 22 yo male, MVC, ETA 8 min. Patient was struggling with EMT [Emergency Medical Technician] providers while being wheeled in. Patient had splints in place at left ankle and right forearm.
> EMT reported: "22 yo male in MVC, 5 minute extrication, who started seizing enroute.
> TA [Trauma Attending] called out for transfer to trauma bay stretcher while report ongoing, as patient was still struggling and staff was working hard to control patient. There was no count for the move, EMT asked "is everyone ready? Okay go" and patient was moved. TA called out check that the bed was locked.
> TL: "what was your transit time?"
> EMT: "about 20 minutes. He has no past medical history. We gave him 5 of versed"
> TA: "secure the IV, get the first blood pressure and prepare to intubate"
> Throughout all of the above, it continued to take 4 people to hold the patient down.

There are multiple issues involved in this case, and they represent one clear example of the difficulty involved in the consideration of the team members' behavior and the false utility of the concept of error. Any evaluation of the behavior forces the question "Should they have done 'X'?", where 'X' could be any of the following: TA waiting for report to finish prior to calling for transfer; formal count to ensure everyone ready prior to move; placing cervical collar to protect cervical spine; etc. Indeed, a case could be made for any of those as "correct" behavior, and yet the context of the situation created behavior completely different from those expectations. If we cannot say they should have done 'X', we can certainly say that 'X' is in the range of expected behaviors. We can say the same, however, of the observed behaviors. Our only recourse, therefore, is to describe the events and the execution of tasks. Consider the team behaviors, then as *task* *adaptations*, rather than binary <error-not error> events. ALL tasks – mental or physical – are adapted minute-to-second-to-nanosecond (e.g. within a person to govern the strength of muscle contraction, the preparatory secretion of stress hormones, the depth of inspiration to acquire more oxygen, etc) (Sapolsky, 1994; Pulakos, Arad, Donovan, & Plamondon, 2000). The cognitive moments of seeing a thrashing, combative patient has meaning in the setting of trauma that is not present at all if someone is in a different scenario. We cannot say it was "wrong" to immediately intubate this patient, nor that it would have been "right" to attempt to sedate them with medication or even conversation first. We can say that this instantiation of a trauma resuscitation rapidly protected the patient's airway (a primary concern for every trauma), permitted diagnostic work-up to proceed quickly, and (though not described in this short section) permitted rapid treatment of a few essential and time dependent injuries.

Klinect accounts for this type of event in his analysis by permitting observers to code protocol violations with "intentional noncompliance" – a sort of error-with-an-asterisk-attached[7]. Though this sounds potentially damning, he means it simply as "the actor made a conscious decision to do 'a' instead of the expected 'b'". Rather than a condemnation, it is transcribed as a simple observation. Even if that is the case, how does knowing this help the individual system components, or even the system as a whole? It can only help as we understand the context of the decision, and when framed as a question of how the system created the behavior as the best option for the actor *from his point of view* (Woods, Dekker, Cook, & Johannesen, 2010).

Health care adds another kink in the armor for TEM as an observational filter. Klinect's model asks us to accept that all undesired states are the result of team error. I cannot speak for the cockpit, but if we are asking how safe we can become while caring for patients, and how much risk we can avoid, Undesired States must include events that are known to be life-threatening, only occur because of intervention, and still are the result of absolutely sound medical practice. Consider hemostasis (appropriate clotting of blood) as an example. There is no debate surrounding the idea that controlling hemorrhage is paramount in avoiding major complications and death during trauma resuscitations. Once a trauma patient is admitted to the hospital and hemorrhage is controlled, however, they are at risk for clotting in their major veins,

[7] This event was coded exactly that way – as intentional noncompliance in stabilizing cervical spine. From one point of view, the only way to achieve airway protection and spinal stabilization was to take control of the patient through intubation. Intentional noncompliance becomes the way to document a workaround.

subsequent pulmonary embolism, and death. Protocols exist, as they should, for prophylaxis against such complications, but the effective protocols involve administration of blood thinners. Blood thinners can cause bleeding, and we occasionally see subsequent massive bleeding in patients caused directly by our own best efforts to prevent fatal pulmonary emboli. Clearly any secondary hemorrhage due to blood thinners is an Undesired State. It did not arise from team error, however, but from perfect team performance. Any analysis of risk creation, therefore, must not limit itself only to behavior that is outside the bounds of Work as Imagined; it must expand to include how each agent adapted to the situation as it unfolded.

Replacing 'error' with 'adapted task' is completely consistent with the initial goals of LOSA. The conceptualization of the LOSA isolated observations of individual behavior from receiving feedback or task specific learning specifically to achieve as close-to-normal behavior as possible in the cockpit. This allows us to learn qualitative information about the system. Suddenly we have the capacity for tremendous power, just as Helmreich and Klinect intended (Helmriech, Klinect, & Wilhelm, 1998). We can now ask two crucial questions. First, how accurate, or even useful, is work as imagined? Second, how successful is work as done? These are fundamentally different and equally important questions. They are nevertheless regularly blurred in our discussions of work simply because the concept of error clouds our vision. Each of these questions deserves a separate answer.

HOW SUCCESSFUL IS WORK AS DONE?

Klinect used the metaphor of defensive driving as a useful way to think about threat and error management within the cockpit. His metaphor rests on considering the key to pilot success to be anticipation of potential trouble, recognition of problems (and potential problems) as soon as they arise, and effective management of sudden risk (Merritt & Klinect, 2006). In this model we are primarily concerned not with the problem of error, nor with the problem of threats, but with the process of work. This process, irrespective of our expertise, training, or system design, will include both intuitive and deliberative performance on the part of the humans involved (Railton, 2016). Irrespective of the mode of the behavior, we must be concerned with these efforts as process rather than outcome. Let us consider an example: If I get in my car and drive to the hospital, and no-one is run off the road, and I don't get in an accident, and arrive at the hospital on time, then I am successful. We all know that a large portion of that drive will occur outside my consciousness, operating in Kahneman's System 1 mode (Kahneman, 2011). In that mode I will make many small adjustments to the curves, the rates of speed in the other cars on the road. I might drive with the radio on, or with or without a GPS. I might drive with one or the other hand on the wheel, or have my hands in the 10 and 2 o'clock position or the 5 and 7 o'clock position. Any hand position I could have on the wheel could only be considered a task adaption rather than an error, as I would have violated no driving laws. Good systematic observation, however, would note how often my hands changed position, where I tended to have them, and how I responded to road conditions in each of the various positions. It would also note the presence of GPS, my gazing at the sunrise or details I passed by, and what radio station was on, and perhaps whether the station played music or was news. All of these details create context for my behavior. Finally, it would also note when I braked as I approached cars, the speed at which I took curves, and it would integrate the other data re: radio, hand position, etc., to create a narrative of my drive. Followed in a prospective fashion, many levels of detail will be lost, but some behaviors, such as those I mentioned and likely countless more you could imagine, will be big enough to catch the eye of the observer.

Moving from the normatively based observation confined to errors to this more broadly contextual approach is extremely liberating. First, there need be no judgement involved in something like hand position on the steering wheel. Second, it opens a much larger field of behavior for analysis. If the concept of error only asks us to get over a hypothetical performance bar, the concept of task adaptation permits the sky to be the limit in terms of performance. Much of the behavior will be noise rather than signal, and it requires careful analysis to see associative patterns (e.g. the inverse association of patient injury severity and likelihood of following spinal precautions in this series). These cannot be considered causal, but they certainly can shape discussion within the system as it learns how it performs.

Such optimism regarding the abandonment of normative behavior as a measurement, however, should not cloud our eyes to the fact that some behavior may be contextually egregious because of the extra-ordinary high risk it creates – e.g. choosing to drive on the left side of the road in the Unites States, or under intoxication. Some behaviors, though they may be correctly classified as task adaptations, can be considered wide enough deviations that the system is forced to consider them egregious violations. These egregious examples, however, are not considered error either, as in each case they have proceeded to the level of illegal.[8]

What does this mean for our analysis of observed performance? To answer our question "How successful is Work as Done?", we cannot simply look at outcome, because it tells us nothing of the cost in terms or resources or effort. Sustainability mandates relative efficiency and safety, and we (as a System) must know the costs of that performance freedom. One type of example was offered with respect to team resilience in the face of inadequate resources. Another prime example is how well procedures are followed when acuity is low. The consideration of the failure to follow protocol in the high acuity cases is only meaningful when considered in the context of how reliably the cervical collar was placed during the low acuity cases. Suppose for a moment, I returned the data on failing to apply the collar 12% of the time to the trauma service as a simple statistic. It is not hard to imagine the wasted training time spent to remind most teams of the need for the collar, including the data regarding the times occult injuries were missed. No-one pays attention to such training because they already have that knowledge, and we all know the outcome of such trainings: they typically fail. This series offers a clue as to why: since the system reliably follows the protocol when there is no perceived clinical reason to do so, and only deviates in the settings where the risk is actually higher, the problem isn't lack of desire or understanding of the importance for following "the rules". The "problem", if there is one, is that those rules exist outside the bounds of specific diagnosis or treatment, and we can therefore become effectively blind to them when specific problem solving begins (Kahneman, 2011).

HOW USEFUL IS WORK AS IMAGINED?

The goals for system safety must always be partially contradictory: move the mean performance to a desired setting and control the standard deviation of this performance, but maintain the capability for variability in response to system stress (Rochlin, 1999; Bergstrom, van Winsen, & Henriqson, 2015; Woods, 2015). Deviation from expected performance should be the norm in nearly every systematic observation of a biological system at work (Gould, 1996). One goal of systematic observation, therefore, must always be to modify both Work as Imagined, and Work as Done. Work as Imagined is useful only as a starting point for

[8] In each of those descriptions, from the vantage point of Threat Management and Task Adaptation, the driver has progressed beyond simple task adaptation and moved to the realm of Undesired State. The adapted task preceded the more egregious violation (e.g. putting the car into gear while intoxicated).

us to learn what we (those inside and outside) don't understand about the system. If we observe work being done only to constrain it to the limits of Work as Imagined, we will fail to see emerging threats over time, just as we will fail to see adaptive successes. The question must always be: what have we learned about the current reality of Work as Imagined, and how should it be modified to optimize system performance? The tension between these two is most visible in concepts such as procedural drift or normalization of deviance (Snook, 2000; Vaughn, 1996). In each case the problem was not the presence of system change, but failure to understand it as present (Woods, 2005). Rather than use systematic observation to ask how we need to fix component performance, we can ask how we need to modify the system to optimize component performance. This perspective shift changes how we think about observation in subtle but fundamental ways. Nothing is inherently wrong with measuring Work as Done against Work as Imagined, but it can be done more beneficially without invoking the concept of error. As the sole observer I am aware of the classification of each "error" coded, as well as the round table discussions held regarding each of them. All events coded as error could easily have been considered as task adaptations as this term encompasses Reason's slips, lapses and mistakes as well as concepts such as "intentional noncompliance". It also includes examples such as the example below:

[As background, this patient is elderly, was intubated in the field because of unresponsiveness, was hemodynamically unstable and had exsanguinating hemorrhage from an open fracture of her left leg, for which a tourniquet has been placed. The team has been working to stabilize her for about 40 minutes prior to this section of the narrative. The numbers on the left starting with 1756 are military time designations.]

> 1756 TA asked for verification of resuscitation fluids to date. PrimaryRN answered "4, 4 and 4 plus 4 liters of crystalloid. TA asked for another 500cc bolus. Team is then waiting as foley catheter is inserted.
>
> 1759 Foley in.
>
> 1800 Repeat ABG drawn. BP 127/62, HR 89, saturation 100%.
> ScribeRN asked for vent settings. PrimaryRN answered "AC/17/500/100%/PEEP 5 and LR 200/hr"
>
> 1804 PrimaryRN checked foley catheter and said "Temp 93.9"
>
> 1805 EDA [Emergency Department Attending] noted patient blinking eyes. Assisting SR [Senior Resident] noted patient moving and suggested as an aside: "they should get 100mcg of fentanyl". No one acknowledged, nor did SR pursue suggestion further.
>
> 1806 PrimaryRN asked if team was done with rapid transfuser. TA answered "no". ABG returned with lactate 6.7
>
> 1808 TA announced to deflate lower extremity tourniquet and it was done. TA called for 4 more units to be delivered at 1:1.

> 1809 Temp now 94.8. BP 112/48 and HR 85. TA announced "the pressures are okay, we can go to CT". Team began to arrange patient for transfer to CT. Lines from the IV pole holding the arterial line were connected to the IV pole attached to the head of the bed under the direction of the TA.
>
> 1813 A-line reading now difficult to get. "4 more units PRBCs in" called out by PrimaryRN. Team reviewed A-line to examine difficulty. Line had been clamped during movement to the new IV pole. Line was unclamped, and pressures now 143/69. PrimaryRN called out "are foley, a-line and respiratory therapist ready?" RT [Respiratory Therapist]: "ready"
>
> 1816 Patient suddenly sat up on stretcher. BP noted to be 117/51. EDA noticed and with RNs quickly and easily laid patient back to stretcher. TA called out for 2mg versed please". A few moments later TA stated "give 100mcg fentanyl before the versed". Both were administered and patient sedated.
>
> 1817 PrimaryRN: "2 units more FFP in" PrimaryRN asked where the crash cart was. TA inquired as to why. ScribeRN stated "you should get the RSI [Rapid Sequence Intubation kit] from the cart in case you need to paralyze her in CT. TA: "I agree." PrimaryRN returned with RSI. RNs crosschecked foley and line placement, and transported patient to CT scanner.

What do we see here? The answer is, of course, "whatever you look for" (Lundberg, Rollenhagen, & Hollnaged, 2009). First consider one simple detail: the arterial line is inadvertently clamped when changing poles for transport. This is precisely the sort of work Helmriech and Klinect had in mind in developing TEM as a model for normal cockpit work. It is also the kind of work we recognize as ubiquitous in daily life. Furthermore, it is quickly recognized and addressed – in the TEM world the team performed very well here, because they immediately identified a problem they created, diagnosed the cause and fixed it, preventing it from becoming a link to other events.

A more subtle event is the assisting senior resident who noted that since this patient had blinked and moved she might be waking up, they should medicate her. I noted this as the observer because his statement "they should get 100 micrograms of fentanyl" was quietly stated under his breath, as if it was an idea that occurred to him, but as if he was watching the episode unfold on television rather than directly in front of him. Work as Imagined, however, dictates the suggestion should be repeated until acknowledged by another team member and a decision made. I considered this behavior worth noting because all possible spoken words are transcribed, and I was forced to consider it an "error" only because TEM doesn't permit anything else. Though the roundtable agreed this was acceptable referenced against WAI, we all understood the resident's behavior as entirely normal. This is exactly the type of gray area that would most benefit from broadening to something like "task adaptation". The behavior may have been poorly executed, but the term "error" seems inarticulate at best, and overly damning at worst. Interestingly, I considered this notable behavior to follow not because of his judgment, but because it wasn't the communication technique we teach. Had he repeated the suggestion, gotten acknowledgement from the Trauma Attending his task would have been completed. If the Trauma Attending had considered his request but decided against it, there would also not have been an error – it is an acceptable judgement call to make. This is especially so in this case because the team has been working so hard to control the patient's hypotension. Dosing the patient for sedation could be expected to drop her blood pressure, and waiting for the patient to move a bit more might be exactly what an experienced physician would do. Despite that, during Data Cleaning all participants agreed that the patient sitting upright was an Undesired State, and therein lies the crux of the problem. We can easily agree on Undesired States, but then we treat the behavior that gets us there very differently (Bosk, 1979; Marx, 2001). We quickly become appropriately uncomfortable with the term error because it only describes the orientation of the observer, never that of the participant. Task Adaptation, on the other hand, should work from the same place that we teach teams to debrief. This involves three questions, none of which involve error: What went well? What would you like to do differently? What will you do to make that change? Note that there is nothing preventing the answers from the first two questions from being identical. The example from the narrative above illustrates this perfectly. We could just as easily say the visiting senior resident did exceptional work by showing up to the activation for which he was not required to offer assistance (especially in the context of routinely understaffed teams). We could say he did an excellent job noting the patient moving and advocating for more pain medication. And even he might finally suggest a way he would like to do the same behaviors differently in the future, none of which requires or is even helped by the concept of error. Furthermore, he might have simply been thinking out loud, and decided not to pursue his suggestion because he realized the blood pressure might drop and this would be counterproductive, thus effectively censoring his own verbalizations as he chose to learn while watching the process unfold. We simply can't know – and probably even he can rarely know for sure – how all those ideas played out in his head.

STUDY LIMITATIONS

Discussion of the specific results of a study such as this is extremely difficult because of the sensitivity with which the data much be treated. The success of the entire process is based entirely on the trust of the observed, which includes the institution. Only the most general findings have been able to be shared, and hopefully they have been sufficient to spur interest in this type of work. The short segments of narrative are consistent with those used in LOSA training, and even found as examples on the FAA circular describing the LOSA process (FAA, 2006).

The limitations of this pilot study largely relate to its small sample and the inherent difficulty with systematic observation. Sample size may have limited analysis of some types of threats or errors. Recommendations for LOSA are a minimum of 50 observations, and it appears that 50 to 100 observations would also be best for COSA (Klinect, 2005). Nevertheless, some clear patterns were able to emerge even at this small number of observations. We believe we have actionable information for targeted improvement based on the data collected. Furthermore, we are confident in the utility of COSA to accurately delineate creation and mitigation of risk in vivo.

The major liability of systematic observation is related to bias. Every observer will develop a patterned way of watching the flow of an event, and it this will confer both advantages and disadvantages. Every effort was made to attend to all spoken dialogue within the activation and follow responses. Observations were also systematically organized to record sequence of events and timing of milestone events such as team member arrival, checklist usage, and time to all diagnostic measures throughout the activation. This helped keep the observations focused and sufficiently broad. Use of s single observer in this case offers both the advantage of internal observational consistency, as well as greatly increasing espertise, but at the expense of a singular point of view. Bias toward specific errors, threats or management was addressed through the multidisciplinary data cleaning round tables. LOSA typically discards roughly 20% of errors, and typically adds errors to about 20% of the observations (Klinect-personal communication). No errors were discarded from our observations by the Round Table participants, however, though several were added from review of the narratives. The consensus among the round table participants was that the narratives were objective and clear. Nevertheless, multiple observers may offer significantly different perspectives, and it is probably beneficial to have at least a few different observers.

Systematic observation relies on observing behavior approximating normal operations as closely as possible. LOSA operational characteristics have been clearly defined and were closely followed. Our data from questioning staff on their opinions of team behaviors indicates that we probably observed very close to normal behavior. From 30 observations of between 6-19 people present, only two people queried felt behavior had changed due to the observation. The first was a nurse, though she was overheard voicing her opinion and countered immediately by a second nurse in the room who emphatically stated the opposite. The second was a surgeon who admitted that he wouldn't have worn a protective gown if the observation hadn't been occurring. Finally, since I am regularly in the trauma bay as part of my normal work flow, I can attest to the fact that what I saw during observations was typical of what I see and hear when I'm working alongside the staff.

CONCLUSION

We are at a similar time in the development of safety in medical practice as aviation was in the mid-1990s; established teamwork skills and clinical treatment guidelines exist to guide the process of care, yet an unsatisfactory level of harm still occurs (Makary & Daniel, 2016). Furthermore, clinicians lack an integrated understanding of teamwork as an essential practice in the care of the patient (Carney, West, Neily, Mills, & Bagian, 2010). Though they may receive some teamwork training, nontechnical skills required in teamwork most often remain abstract concepts without clear utility in everyday practice. We tend to become like our patients, expecting to need the "medicine" of teamwork skills only in a crisis, and not understanding that only daily practice keeps our systems healthy (Greenberg, et al., 2007). We must integrate teamwork skills as usable tools in the way clinicians intuitively go about their work (King & Harden, 2013). It is a striking finding that ATLS algorithms in and of themselves do not address teamwork concepts. The algorithms are usable from a clinical care standpoint, but they leave out goals and information sharing methods known to be critical to effective work (Endsley, 1995; Bergrström, Dahlström, Henriqson, & Dekker, 2010).

The literature written by pilots when they were first exposed to Threat and Error Management makes very clear that they understood teamwork differently in light of TEM. Suddenly teamwork skills were tools to help them manage the threats and errors they encountered effectively, rather than "charm school" (Tullo, 2010). TEM became so fundamental to the work of pilots it is now taught in the earliest phases of flight training. Most physicians and staff I have encountered intuitively understand TEM and Task Adaptation in similar fashion.

One of the great values of TEM-based COSA is not only to detect normal work flow, but to see what situations tend to overwhelm teams and how they adapt to the situations they are given using necessary efficiency-thoroughness tradeoffs. Honest evaluation of the teams' work flow will inevitably show that some of these tradeoffs are a wise use of resources given the situation. A method that professionals trust as reflective of their own work flow (whether they lead to failures or successes) can promote staff engagement in learning about system, team, and personal function. The operational characteristics built into LOSA, and adapted as strictly as possible by us, successfully created that environment. We believe that our data has provided us with clear and achievable targets for improvement in our teamwork processes. Most importantly, our teams found the process to be both enlightening and empowering in terms of self-directing their own efforts to change behavior. This work is fundamental to our psychology and to navigating all activities with success not only as evaluated in terms of outcome, but also in terms of the human experiences of purpose, joy and meaning (Barrett, 2017; Conklin, 2016; Frush, Leonard, & Frankel Allan, 2013; Leape, et al., 2012). Though specifically employed in Trauma Activations during this pilot study, we believe this work is adaptable in a myriad of clinical situations, from the procedural suite to the intensive care unit.

Despite the inherent value perceived in systematic observation, as well as the utility of TEM as a starting point, it is clear that the concept of error is more limiting than useful, and actually erodes contextual understanding of events. Considering all tasks – whether cognitive or physical – in terms of their adaptive utility is richer and more meaningful, and it this analysis I propose to explore in the future.

ACKNOWLEDGEMENTS

There are several people without whom this project would have proved even more daunting and certainly either it, or I, would have emerged appearing completely different. In chronological order: first, thanks to James Klinect, and Hawaiian Airlines for agreeing to have an interested surgeon attend and observer training session, and for ongoing discussion during and after the project. Thanks to John Cooley for flying to Hawaii and sharing my interest in this work, and Lynn Hubert for being willing to learn and observe in the initial pilot series. Thanks to Azlyn Goff and Alexey Abramov, for their efforts in assisting with organizing data and building the database. Next, thanks to the trauma service for bravely trusting this process and me personally. Special thanks to the Director of Trauma, Corrado Marini, and the quality team under Renee Garrick for unwavering support.

Thanks to Steve Montague for suggesting the Lund program; friends like that are a great blessing. Lund proved to be the perfect place to thoughtfully consider the implications of the observation process and its results. Throughout the Masters program both Johan Bergström and Anthony Smoker remained the perfect mentors. Lund is unbelievably lucky to have them. Extra thanks to JB for agreeing to guide me through this thesis, especially after having to read so much of my writing in Course 1.

Finally, thanks to my wife Gina, and my daughters Eliza and Sasha, for alternately permitting me to ignore them while working, and accept listening to me ramble about my ideas, as well as my new love for Lund and Sweden.

REFERENCES

Amalberti, R. (2013). *Navigating Safety: Necessary Compromises and Trade-offs - Theory and Practice.* Dordrecht: Springer.

ATLS. (2012). *Advanced Trauma Life Support Student Manual.* Chicago: American College of Surgeons.

Barrett, L. F. (2017). *How emotions are made.* New York : Houghton Mifflin Harcourt.

Bergrstrom, J., Dahlstrom, N., Henriqson, E., & Dekker, S. (2010). Team coordination in escalating situations: An empirical study using mid-fidelity simulation. *Journal of Contingencies and Crisis Management, 18*(4), 220-230.

Bergstrom, J., van Winsen, R., & Henriqson, E. (2015). On the rationale of resilience in the domain of safety: a literature review. *Reliability Engineering and System Safety*, 131-141. Retrieved from http://dx.doi.org/10.1016/j.ress.2015.03.008i

Bosk, C. L. (1979). *Forgive and remember - managing medical failure.* Chicago, IL: The Univeristy of Chicago Press.

Braithwaite, J., Wears, R. L., & Hollnagel, E. (2015). Resilient health care: Turning patient safety on its head. *International Journal for Quality in Health Care, 27*(5), 418-420.

Branlat, M., & Woods, D. D. (2010). How do Systems Manage Their Adaptive Capacity to Successfully Handly Disruptions? A Resilience Engineering Perspective. *Complex Adaptive Systems - Resilience, Robustness and Evolvability: Papers from the AAAI Fall Symposium* (pp. 26-34). Association for the Advancement of Artificial Intelligence.

Carney, B. T., West, P., Neily, J., Mills, P. D., & Bagian, J. P. (2010). Differences in nurse and surgeon perceptions of teamwork: implications for the use of a briefing checklist in the OR. *Journal of American Operating Room Nurse, 91*(6), 722-729.

Cassel, C. K., & Reuben, D. B. (2011). Specialization, subspecialization, and subsubspecialization in internal medicine. *New England Journal of Medicine, 364*(2), 1169-1173.

Conklin, T. (2016). *Better questions: An applied approach to operational learning.* Boca Raton, Florida: CRC Press.

Dekker, S. (2015). The danger of losing situational awareness. *Cognitive Technology and Work*, 159-161.

Dekker, S., & Hollnagel, E. (2004). Human factors and folk models. *Cognitive Technology and Work*, 79-86.

Endsley, M. (2015). Situation awareness: operationally necessary and scientifically grounded. *Cogntive Technology and Work*, 163-167.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Engestrom, Y. (2000). Activity theory as a framework for analyzing and redisigning work. *Ergonomics*, 43(7) 960-974.

Ertel, W. K., & Kellam, J. F. (2015). General assessment and management of the polytrauma patient. In M. TIle, D. L. Helfet, J. F. Kellam, & M. Vrahas, *Fractures of the Pelvis and Acetabulum: Principles and Methods of Management, 4th Ed* (pp. 61-82). Davos Platz: AO Foundation.

FAA. (2006). *Advisory circular: Line operation safety audits.* Washintgon, D.C.: Federal Aviation Administration.

Frush, K., Leonard, M., & Frankel Allan. (2013). Effective teamwork and communication. In M. Leonard, A. Frankel, F. Frederico, K. Frush, & C. Haraden, *The essential guide for patient safety officers, 2nd Edition* (pp. 53-68). Oakbrook Terrace, Illinois: Joint Commission Resources.

Frush, K., Maynard, L., Koeble, C., & Schwendimann, R. (2013). Regulating and monitoring teamwork in healthcare: Issues and challenges. In K. Frush, & E. Salas, *Improving Patient Safety Through Teamwork and Team Training* (pp. 94-104). New York: Oxford University Press.

Gawande, A. (2011, October 3). Personal Best. *The New Yorker.* Retrieved from https://www.newyorker.com/magazine/2011/10/03/personal-best

Gould, S. J. (1996). The Bare Bones of Natural Selection. In S. J. Gould, *Full House: The Spread of Excellence from Plato to Darwin* (pp. 135-146). New York: Harmony Books.

Greenberg, C. C., Regenbogen, S. E., Studdert, D. M., Lipsitz, S. R., Rogers, S. O., Zinner, M. J., & Gawande, A. A. (2007). Patterns of communication breakdowns resulting in injury to surgical patients. *Journal of the Amerian College of Surgeons, 204*(4), 533-540.

Hansson, T. (2015). *Contemporary approaches to activity theory.* Hershey, PA: IGI Global.

Helmriech, R. L., Klinect, J. R., & Wilhelm, J. A. (1998). Models of Threat, Error and CRM in Flight Operations. Retrieved September 20, 2013, from http://flightsafety.org/archives-and-resources/threat-and-error-managment-tem

Hollnagel, E., & Amalberti, R. (2001). The Emperor's New Clothes or Whatever Happened to "Human Error"? *Proceedings of the 3rd Conference on Human Error and System Safety Development.* Linkoping, Sweden.

Jauhar, S. (2014, August 9). One patient, too many doctors: The terrible expense of overspecialization. *Time.* Retrieved January 15, 2018, from http://time.com/3138561/specialist-doctors-high-cost/

Jordan, G. L. (1985). The impact of specialization on health care. *Annals of Surgery, 201*(5), 537-544.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar,Straus and Giroux.

Kalkwarf, K. J., & Cotton, B. A. (2017, December). Resuscitation for hypovolemic shock. *Surgical Cinics of North America, 97*(6), 1307-1321.

Kant, V. (2016). Supporting the human life-raft in confrontin the juggernaut of technology: Jens Rasmussen 1961-1986. *Applied Ergonomics*, 1-11.

Klinect, J. R. (2005). *Line operations safety audit: A cockpit observation methodology for monitoring commercial airline safety performance.* Austin, TX: The University of Texas at Austin.

Klinect, J. R., Murray, P., Merritt, A., & Helmriech, R. (2003). Line Operation Safety Audits (LOSA): Definition and Operating Characteristics. *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 663-668). Dayton, OH: The Ohio State University.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (2000). *To err is human: building a safer health system.* Washington, D.C.: Institute of Medicine.

Leape, L. L., Shore, M. F., Dienstag, J. L., Mayer, R. J., Edgman-Levitan, S., Meyer, G. S., & Healy, G. B. (2012). A culture of respect, part 2: Creating a culture of respect. *Academic Medicine, 87*(7), 853-858.

Lundberg, J., Rollenhagen, C., & Hollnaged, E. (2009). What-you-look-for-is-what-you-find: The consequences of underlying accident models in eight accident investigation manuals. *Safety Science, 47*(10), 1297-1311.

Makary, M. A., & Daniel, M. (2016). Medical error - the third leading cause of death in the US. *British Medical Journal, 353*, i2139.

Marx, D. (2001). *Patient Safety and the "Just Culture": A Primer for Health Care Executives.* New York, NY: Columbia University.

Maurino, D. (2005). *Threat and error management* . Vancouver, BC: Canadian Aviation Safety Seminar. Retrieved from http://flightsafety.org/files/maurino.doc

Merritt, A., & Klinect, J. (2006). *Defensive Driving for Pilots: An introduction to Threat and Error Management.* Austin: The LOSA Collaborative. Retrieved January 5, 2013, from https://flightsafety.org/files/tem_dspt_12-6-06.pdf

Perry, S. J., & Wears, R. L. (2012). Underground adaptations: case studies from health care. *Cognitive Technology and Work, 14*, 253-260.

Perry, S. J., Wears, R. L., & McDonald, S. S. (2013). Implementing team training in the emergency department: The good, the bad, the unexpected and the problematic. In K. Frush, & E. Salas, *Improving Patient Safety Through Teamwork and Team Training* (pp. 129-135). New York: Oxford University Press.

Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology, 85*(4), 612-624.

Railton, P. (2016). Introducing Homo prospectus. In M. E. Seligman, P. Railton, R. F. Baumeister, & C. Sripada, *Homo prospectus* (pp. 7-31). New York: Oxford University Press.

Rasmussen, J. (1988). Human error mechanisms in complex work environments. *Reliability Engineering and System Safety, 22*, 155-167.

Rasmussen, J. (1997). Risk Management in a Dynamic Society: A Modelling Problem. *Safety Science, 2*(2/3), 183-213.

Reason, J. (1990). *Human Error.* New York: Cambridge University Press.

Reason, J. (2008). *The Human Contribution: Unsafe Acts, Accidents and Heroic Recoveries.* Surrey, England: Ashgate Publishing Company.

Rochlin, G. I., La Porte, T. R., & Roberts, K. H. (1987, Autumn). The Self-Designing High-Reliability Organization: Aircraft Carrier Flight Operations at Sea. *Naval War College Review*, pp. 76-90.

Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation Awareness in Team Perfromance: Indications for Measurement and Training. *Human Factors, 37*(1), 123-136.

Salmon, P. M., Walker, G. H., & Stanton, N. A. (2015). Broken components versus broken systems: why it is systems not people that lose situational awareness. *Cognitive Technology and Work, 17*(2), 179-183.

Sapolsky, R. M. (1994). *Why Zebras Don't Get Ulcers.* New York: Henry Holt & Co.

Scott, S. D., Hirschinger, L. E., Cox, K. R., Brandt, J., & Hall, L. W. (2009). The natural history of recovery for the healthcare provider "second victim" after adverse patient events. *Quality and Safety in Health Care, 18*, 325-330.

Seligman, M. E., Railton, P., Baumeister, R. F., & Sripada, C. (2016). *Homo prospectus.* New York: Oxford University Press.

The Joint Commission. (2008, July 9). *Behaviors that undermine a culture of safety.* https://www.jointcommission.org/assets/1/18/SEA_40.PDF: The Joint Commission.

The Joint Commission. (2016). *Comprehensive certification manual for disease-specific care.* https://www.jointcommission.org/assets/1/6/DSC_05_SE_update_CURRENT.pdf: The Joint Commission.

Tullo, F. J. (2010). Teamwork and organizational factors. In B. G. Kanki, R. L. Helmreich, & J. Anca, *Crew Resource management, 2nd Ed.* (pp. 59-78). San Diego, California: Elsevier.

Woods, D. D. (1988). Coping with Complexity: The psychology of human behavior in complex systems. In L. P. Goodstein, H. B. Andersen, & S. E. Olsen, *Tasks, Errors and Mental Models* (pp. 128-148). London: Taylor and Francis.

Woods, D. D. (2015). Four concepts for resilience and the implications for the future of resilience engineering. *Reliability Engineering and System Safety*, 5-9.

Woods, D. D., & Hollnagel, E. (2006). *Joint Cognitive Systems: Pattens in Cognitive Systems Engineering.* Boca Raton: Taylor & Francis Group.

Woods, D. D., Dekker, S., Cook, R., & Johannesen, L. (2010). *Behind Human Error.* Burlington: Ashgate.

Riskhantering och samhällssäkerhet
Lunds tekniska högskola
Lunds universitet
Box 118
221 00 Lund

http://www.risk.lth.se

Division of Risk Management and Societal Safety
Faculty of Engineering
Lund University
P.O. Box 118
SE-221 00 Lund
Sweden